



**PHD**

**Search Satisfaction**

**Choice Overload, Variety Seeking and Serendipity in Search Engine Use**

Chiravirakul, Pawitra

*Award date:*  
2015

*Awarding institution:*  
University of Bath

[Link to publication](#)

**Alternative formats**

If you require this document in an alternative format, please contact:  
[openaccess@bath.ac.uk](mailto:openaccess@bath.ac.uk)

Copyright of this thesis rests with the author. Access is subject to the above licence, if given. If no licence is specified above, original content in this thesis is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC-ND 4.0) Licence (<https://creativecommons.org/licenses/by-nc-nd/4.0/>). Any third-party copyright material present remains the property of its respective owner(s) and is licensed under its existing terms.

**Take down policy**

If you consider content within Bath's Research Portal to be in breach of UK law, please contact: [openaccess@bath.ac.uk](mailto:openaccess@bath.ac.uk) with the details. Your claim will be investigated and, where appropriate, the item will be removed from public view as soon as possible.

# **Search Satisfaction: Choice Overload, Variety Seeking and Serendipity in Search Engine Use**

**Pawitra Chiravirakul**

A thesis submitted for the degree of Doctor of Philosophy

University of Bath

Department of Computer Science

October 2014

## **COPYRIGHT**

Attention is drawn to the fact that copyright of this thesis rests with the author. A copy of this thesis has been supplied on condition that anyone who consults it is understood to recognise that its copyright rests with the author and that they must not copy it or use material from it except as permitted by law or with the consent of the author.

This thesis may be made available for consultation within the University Library and may be photocopied or lent to other libraries for the purposes of consultation.

---

Pawitra Chiravirakul



## **ACKNOWLEDGEMENT**

I would like to express deep appreciation to my academic supervisor, Professor Stephen J. Payne, for his continuous support, guidance and encouragement throughout my PhD. This thesis would not exist and I would not have become an effective researcher as I am now without his supervision. I would also like to thank Dr. Leon Watts for his kind advice and thanks to all past and present members of the HCI group in the Department of Computer Science at the University of Bath for their support and friendship.

I am grateful to the Faculty of Information and Communication Technology at Mahidol University for providing me this great opportunity to continue my studies abroad. I have gained the invaluable knowledge and experience in academia and in life in general.

Last but not least, I would like to thank my family, friends and colleagues who have helped and encouraged me throughout my research journey away from home. Thanks for always believing in me.



## **ABSTRACT**

Users of current web search engines are often presented with a large number of returns after submitting a search term and choosing from the list might lead to them suffering from the effect of “choice overload”, as reported in earlier work. However, these search results are typically presented in an ordered list so as to simplify the search process, which may influence search behaviour and moderate the effect of number of choices. In this thesis, the effects of the number of search returns and their ordering on user behaviour and satisfaction are explored.

A mixed methods approach combining multiple data collection and analysis techniques is employed in order to investigate these effects in terms of three specific issues, namely, choice overload in search engine use, variety seeking behaviour in a situation where multiple aspects of search results are required, and the chance of encountering serendipity. The participants were given search tasks and asked to choose from the sets of returns under experimental conditions. The results from the first three experiments revealed that large numbers of search results returned from a search engine tended to be associated with more satisfaction with the selected options when the decision was made without a time limit. In addition, when time was more strongly constrained the choices from a small number of returns led to relatively higher satisfaction than for a large number. Moreover, users’ behaviour was strongly influenced by the ordering of options in that they often looked and selected options presented near the top of the result lists when they perceived the ranking was reliable. The next experiment further investigated the ranking reliance behaviour when potentially useful search results were presented in supplementary lists. The findings showed that when users required a variety of options, they relied less on the ordering and tended to adapt their search strategies to seek variety by browsing more returns through the list, selecting options located further down, and/or choosing the supplementary web pages provided. Finally, with the aim of illustrating how chance encountering can be supported, a model of an automated synonym-enhanced search was developed and employed in a real-world literature search. The results showed that the synonym search was occasionally useful for providing a variety of search results, which in turn increased users’ opportunity to come across serendipitous experiences.



## PUBLICATION

**This thesis has resulted in the following peer-reviewed publication:**

Chiravirakul, P., & Payne, S. J. (2014, April). Choice overload in search engine use?. In *Proceedings of the 32nd annual ACM CHI conference on Human factors in computing systems* (pp. 1285-1294). ACM.





# CONTENTS

<b>Acknowledgement.....</b>	<b>i</b>
<b>Abstract .....</b>	<b>iii</b>
<b>Publication.....</b>	<b>v</b>
<b>List of Figures .....</b>	<b>xi</b>
<b>Chapter 1 Introduction .....</b>	<b>1</b>
1.1 Thesis Overview .....	1
1.2 Definitions.....	4
1.2.1 Search Engine.....	4
1.2.2 Choice .....	4
1.3 Research Questions.....	5
1.4 Research Methods.....	6
1.4.1 Research Ethics .....	6
1.5 Thesis Outline .....	7
<b>Chapter 2 Literature review .....</b>	<b>11</b>
2.1 Chapter Overview .....	11
2.2 Choice Overload .....	13
2.3 Search Engine .....	18
2.3.1 Search Engine Use .....	18
2.3.2 Search Engine Design .....	19
2.3.3 Search Behaviour Models .....	22
2.4 Satisfaction.....	24
2.4.1 Satisfaction Definitions and Dimensions .....	25
2.4.2 Decision Factors and Their Effects on Satisfaction .....	27
2.4.3 Search Satisfaction in a Computer-based Environment.....	31
2.5 Choice Overload for Search Engines .....	32
2.6 Variety Seeking.....	34

## CONTENTS

---

2.7 Serendipity .....	38
2.8 Chapter Summary .....	42
<b>Chapter 3 Choice Overload in Search Engine Use .....</b>	<b>43</b>
3.1 Chapter Overview .....	43
3.2 Introduction.....	45
3.3 Experiment 1: Number of Options and Iteration .....	46
3.3.1 Method .....	46
3.3.2 Results.....	49
3.3.3 Discussion .....	55
3.4 Experiment 2: Number of Options and Ranking .....	56
3.4.1 Method .....	56
3.4.2 Results.....	59
3.4.3 Discussion .....	64
3.5 Experiment 3: Number of Options, Ranking and Time pressure.....	65
3.5.1 Method .....	65
3.5.2 Results.....	68
3.5.3 Discussion .....	74
3.6 General Discussion .....	74
3.6.1 Qualitative Findings.....	76
3.6.2 Limitations and Future Work.....	77
3.7 Chapter Summary .....	77
<b>Chapter 4 Variety Seeking and Chance Encountering .....</b>	<b>79</b>
4.1 Chapter Overview .....	79
4.2 Introduction.....	81
4.3 Experiment 4: Number of Options and Supplementary Web Pages...	86
4.3.1 Method .....	86
4.3.2 Results.....	91
4.3.3 Discussion .....	107
4.3.4 Limitations and Future Work.....	111
4.3 Chapter Summary .....	112
<b>Chapter 5 Serendipity and Literature Search .....</b>	<b>115</b>
5.1 Chapter Overview .....	115
5.2 Introduction.....	117

---

5.3 Study 5: Automated Synonym-enhanced Search for Serendipitous Experience.....	121
5.3.1 Method .....	121
5.3.2 Analysis.....	126
5.3.3 Discussion .....	140
5.3.4 Study Limitations and Future Work.....	142
5.4 Chapter Summary .....	142
<b>Chapter 6 Conclusion .....</b>	<b>145</b>
6.1 Summary Overview .....	145
6.2 Thesis Summary.....	145
6.2.1 A Review of the Thesis Findings .....	145
6.2.2 An Integration of the Thesis Findings.....	149
6.2.3 A Theoretical Sketch of Search Satisfaction Judgements.....	150
6.2 Discussion of the Findings and Contributions .....	152
6.3 Limitations and Future Directions .....	155
6.4 Concluding Remarks.....	157
<b>References.....</b>	<b>159</b>
<b>Appendix A Ethics Checklist .....</b>	<b>175</b>
<b>Appendix B Experiment 1 Materials .....</b>	<b>179</b>
<b>Appendix C Experiment 2 Materials .....</b>	<b>183</b>
<b>Appendix D Experiment 3 Materials .....</b>	<b>187</b>
<b>Appendix E Experiment 4 Materials .....</b>	<b>191</b>
<b>Appendix F Study 5 Materials.....</b>	<b>197</b>



## LIST OF FIGURES

Figure 3.1 Experiment 1: An example of a computer screen presenting six options.....	47
Figure 3.2 Experiment 1: An example of a computer screen presenting twenty-four options (on four SERPs) .....	47
Figure 3.3 Experiment 1: Mean position of the selected web pages (left), and number of SERPs viewed before selecting (right). Error bars show standard errors. ....	49
Figure 3.4 Experiment 1: Mean satisfaction with the selected web pages. Error bars show standard errors.....	51
Figure 3.5 Experiment 1: Mean overall satisfaction for the selected web pages (left), and mean overall satisfaction for the lists of options returned (right). Error bars show standard errors. ....	51
Figure 3.6 Experiment 2: An example of a computer screen presenting twenty-four options (on four SERPs) in the Random ranking condition .....	58
Figure 3.7 Experiment 2: Mean position of the selected web pages (left), and number of SERPs viewed before selecting (right). Error bars show standard errors. ....	60
Figure 3.8 Experiment 2: Mean satisfaction with the selected web pages. Error bars show standard errors.....	61
Figure 3.9 Experiment 2: Mean overall satisfaction for the selected web pages (left), and mean overall satisfaction for the lists of options returned (right). Error bars show standard errors. ....	61
Figure 3.10 Experiment 3: An example of a computer screen presenting twenty-four options (on four SERPs) in the Random ranking condition with a digital clock.....	67

## LIST OF FIGURES

---

Figure 3.11 Experiment 3: Mean position of the selected web pages (left), and number of SERPs viewed before selecting (right). Error bars show standard errors. ....	68
Figure 3.12 Experiment 3: Mean satisfaction with the selected web pages (left), and mean satisfaction with selected web pages in which the data were pruned (right). Error bars show standard errors. ....	69
Figure 3.13 Experiment 3: Mean relevance (left), and confidence (right). Error bars show standard errors. ....	70
Figure 3.14 Experiment 3: Mean overall satisfaction for the selected web pages (left), and mean overall satisfaction for the lists of options returned (right). Error bars show standard errors. ....	71
Figure 4.1 Experiment 4: An example of a computer screen presenting forty-eight options (on eight SERPs) with no supplementary web pages .....	90
Figure 4.2 Experiment 4: An example of a computer screen presenting twenty-four options (on four SERPs) with synonym-supplementary web pages .....	91
Figure 4.3 Experiment 4: An example of a computer screen displaying a selected web page .....	91
Figure 4.4 Experiment 4: Mean number of SERPs viewed before selecting for the single selected target (left) and three selected targets conditions (right). Error bars show standard errors. ....	92
Figure 4.5 Experiment 4: Modal number of SERPs viewed per participant. ....	93
Figure 4.6 Experiment 4: Modal number of SERPs viewed per participant for the single selected target. ....	93
Figure 4.7 Experiment 4: Modal number of SERPs viewed per participant for the multiple selected targets. ....	94
Figure 4.8 Experiment 4: Mean position of the selected web pages in the single selected target condition. Error bars show standard errors. .	95
Figure 4.9 Experiment 4: Mean position of the selected web pages in the multiple selected targets condition. Error bars show standard errors. ..	96

---

Figure 4.10 Experiment 4: Top position of the selected web pages (left), middle position of the selected web pages (middle), and last position of the selected web pages (right). Error bars show standard errors. ....	96
Figure 4.11 Experiment 4: Mean satisfaction with selected webpages independent of the question, for the single selected target (left) and three selected targets conditions (right). Error bars show standard errors....	97
Figure 4.12 Experiment 4: Mean satisfaction with the search and selection process, for the single selected target (left) and three selected targets conditions (right). Error bars show standard errors.....	98
Figure 4.13 Experiment 4: The number of selected supplementary web pages, for the single selected target (left) and three selected targets conditions (right). Error bars show standard errors. ....	99
Figure 4.14 Experiment 4: Mean satisfaction with the supplementary selected web pages concerning the questions. Error bars show standard errors. ....	100
Figure 4.15 Experiment 4: Mean satisfaction with the search and selection process (left), and confidence in selecting the best web pages (right). Error bars show standard errors.....	101
Figure 5.1 Study 5: An example of the computer screen presenting the keyword-search results and the synonym-search results lists with two options.....	123
Figure 5.2 Study 5: An example of the computer screen presenting the keyword-search results and synonym-search results lists with 48 options .....	124
Figure 5.3 Study 5: Descriptive results under the 2-synonym-option set condition .....	128
Figure 5.4 Study 5: Descriptive results under the 48-synonym-option set condition .....	128
Figure 6.1: The effects of the factors manipulated in the thesis studies on search behaviour and outcomes .....	150





# **CHAPTER 1**

## **INTRODUCTION**

### **1.1 THESIS OVERVIEW**

Search engines typically allow access to a vast amount of information from a wide variety of information sources, simply by specifying text that the source must contain. The use of search engines is extremely widespread in everyday and professional life, and presumably allows users to gain valuable knowledge concerning their topic of interest from the provided search results.

Nevertheless, some limitations may restrict the value of search engines. For example, although a large amount of information is in principle available, users can only access a small proportion due to information processing limitations, such as time constraints. Deciding which of the search returns to access is itself a demanding cognitive decision task. Designers of search engines seem to have taken into consideration such constraints by organising search results based on ranking algorithms, whereby the most relevant results come at the top of the list to some extent (e.g. Page et al., 1999). However, there are a limited number of variables that a ranking algorithm can take into account, such as overlap with the specified search terms, the links among web pages, and the structure of web sites, which can impose boundaries on the relevance ordering. In addition, the reliance on a ranking algorithm may reduce the users' awareness of useful information from other sources that are not presented in a search results list. Moreover, in the search process, results returned are generated based on the keywords entered by users. Some important and relevant information that has been indexed by a search engine may not be retrieved if the users form inadequate search terms, perhaps through limited knowledge

of the subject, e.g. they may not be aware of inconsistencies in the terms used for a particular topic of interest.

Some research related to search engines has focused on increasing the efficiency of these systems through an improvement of the ranking algorithms (e.g. Agichtein et al., 2006; Bar - Ilan et al., 2007; Gao et al., 2009). However, little attention has been paid to how the number of search results returned affects the satisfaction of users regarding their search outcomes. Furthermore, although the influence of ranking algorithms on search behaviour has been investigated in many studies in the field of information retrieval, the effects on user satisfaction with the search process and the outcomes have not been addressed. A deeper understanding of the effects of search engine designs on users' satisfaction is important, if the usability of these systems is to be improved in order to provide support not only for a simple search requirement, such as fact-finding or navigational searches, but also for problem solving processes where information gathering is more complex, such as when the users seek information in order to generate a new idea or look for some inspiration for their creativities.

This thesis aims to understand factors that might improve user satisfaction with selected search results for information seeking tasks, including the number of search results returned by a search engine, the ordering of these returns, the possible desire to find variety among search results, and whether serendipitous discoveries can be encouraged by returning results from outside the standard set of returns. In particular, the effect of the number of search results returned and that of their ordering on search behaviour and satisfaction were investigated, which the existing literature in the online information seeking context has not explored in detail. Regarding the fact that search engines typically return so many results, choosing from the list might result in the user suffering from the effect of “choice overload”, and preliminary work has reported just such an effect (Oulasvirta et al., 2009). In this thesis, a series of experiments is conducted to investigate the choice overload effect and its moderators in the context of search engine use. In addition, the role of ordering of search returns as a potential mitigation of choice overload is examined as well as the effects of such a design on users' behaviour. Moreover, further investigation is undertaken to explore “variety seeking” behaviour of the users and their chance of encountering “serendipity” in situations where variety among the search results is indicated or required. More specifically, search behaviour and outcomes when either additional search results from synonym searches or results

---

from further down the results list presented as supplementary options, are examined. In order to illustrate how the implications for search engine design based on a synonym search can be applied to increase the opportunity to encounter serendipitous experiences in search engine use, the model of an automated synonym-enhanced search is developed and assessed in the real-world task of a literature search. To tackle these issues, the studies presented in this thesis involve a combination of quantitative and qualitative research techniques.

The design of search engines should treat usability as a major criterion, e.g. ease of use, navigating and understanding the system interface. Designers of search engines presumably wish to improve the usability regarding the users' search satisfaction in more complex situations and to help them overcome any limitations in cognitive skill or resources. The work presented in this thesis is underpinned by the perspective of user centred design with the aim of contributing a richer understanding of search engine use that might inform design improvements.

The remainder of this chapter is structured as follows: In section 1.2 the terms that are used throughout this thesis are introduced as well as the aspects that will be covered are briefly identified. Section 1.3 defines the research questions that guide the work presented. This then leads to Section 1.4, which explains the research methodology applied to investigate the issues. Finally, an outline of each chapter is presented in Section 1.5, including an outline of each study conducted.

### 1.2 DEFINITIONS

This thesis covers work that is interdisciplinary and hence, although narrow in focus, the potential meanings of the terms used are wide. Consequently, it is necessary to define clearly the terms used throughout the work so as to avoid any misunderstanding regarding their interpretation.

#### 1.2.1 SEARCH ENGINE

A **search engine** is a program that searches for documents in a database, based on specified keywords (the search terms or search query), and returns a list of documents which contain what has been asked for<sup>1</sup>. The term is used most often in this thesis to refer to a web search engine, such as Google, Bing or Yahoo!, which has been designed to search for pages on the World Wide Web. The results returned from a search-engine search are generally ordered by a **ranking algorithm** that ranks web pages based on a set of criteria, which is not typically available to the users. These returns are presented in the form of pages called **search engine results pages (SERPs)**, with each result item normally including: title, a reference to the original document, and a short description called a **snippet** that shows a few sentences or phrases from the document's content that match with the user's search terms. For this thesis, search results are regarded as options to be chosen among in a decision making process and one aim is to explore users' satisfaction with their selected search result. In addition, the focus is on search tasks that are open-ended, that is, when the correctness of search results cannot be assessed by a definite judgement (i.e., yes or no) the way it can when searching for a fact or a known target.

#### 1.2.2 CHOICE

**Choice** refers to a decision to select one or more options on the basis of judged quality or worthiness of those available<sup>2</sup>. **Choice overload** pertains to a situation in which the large number of options to choose from negatively affects behaviour and responses of decision makers or consumers, such that there is: avoidance of choice, reduced consumption and/or lower satisfaction (Iyengar & Lepper, 2000; Schwartz, 2004). The impact on satisfaction is a particular concern of this thesis. A **consideration set** is a

---

<sup>1</sup> Source: Oxford Dictionary, <http://www.oxforddictionaries.com/definition/english/search-engine>

<sup>2</sup> Source: Oxford Dictionary, <http://www.oxforddictionaries.com/definition/english/choice>

---

subset of options that decision makers filter and assess in detail when making a decision in order to simplify the process (Narayana & Markin; 1975). In the context of search engine use, the consideration set may be based on SERPs or subjectively defined by the search engine user and in the experiments reported in this thesis the number of SERPs viewed is considered to be representation of the searchers' consideration set.

### 1.3 RESEARCH QUESTIONS

The work described in this thesis is guided by the following research questions.

The first research question is: **RQ1: Does choice overload affect users' satisfaction with search engine use?** This question is explored, firstly, by investigating the effect of choice overload in a controlled experimental setting environment, where the number of options returned by a search engine was manipulated. Secondly, some potential moderators of the choice overload effect are examined. The experiments reported in Chapter 3 present the findings that addressed this research question.

The second research question addressed is: **RQ2: What are the effects of the ordering of search results on the behaviour and satisfaction of search engine users?** The work reported in Chapter 3 partly examines the role of the ordering of search returns with regards to the extent to which it can mitigate the choice overload effect.

Next, in Chapter 4 a task is introduced which encourages seeking variety among search results returned. The effects on the users' behaviour and satisfaction are investigated in two situations: single option or multiple options selection. This is aimed at answering the third research question: **RQ3: Does a search task which encourages users to seek various targets affect users' behaviour and satisfaction in search engine use?** In addition, users' behaviour is explored in terms of variety seeking and chance encountering.

Finally, the study described in Chapter 5 explores the fourth research question: **RQ4: How can an automated synonym-enhanced search support a literature search process and increase the chance of encountering a serendipitous experience?** This is undertaken using a qualitative approach, with the benefits and limitations of search results generated by a synonym search term and users' requirements in a literature search being examined.

### 1.4 RESEARCH METHODS

As previously stated, the research presented in this thesis focuses on search engine users' satisfaction and behaviour, with an empirical approach being adopted to investigate these issues. Additionally, the issues explored here relate to both the human and technical aspects, which I hope will have implications for search engine designs mainly in the context of understanding search behaviour and outcomes. Hence, a mixed method approach combining multiple data collection and analysis techniques is employed. The rationale for mixing methods is that quantitative and qualitative approaches address the issues from different perspectives and thus, are complementary, thereby allowing for more complete analysis of the subject matter.

For each research question, approaches that are appropriate for addressing it are employed. More specifically, for each study, the methods used are identified by considering their suitability in examining the particular issues of interest. Throughout this thesis, Likert-type scales (Likert, 1932; Dawes, 2008; Cummins, & Gullone, 2000) are used to collect quantitative data and open-ended questionnaires are employed to gather qualitative responses. In addition, semi-structured interviews are used in the study reported in Chapter 5 for obtaining insights about individual perceptions on the use of search engines, as the utility of this method has been illustrated by other researchers (Reneker, 1993; Rieh, 2004; Vandenbosch & Huff, 1997). Furthermore, in much of the extant research that has studied information-seeking behaviour, Google has been used as a research tool to investigate interaction activities between users and the search engine. Consequently, it is deemed appropriate that Google is employed as the information retrieval system to generate the collection of materials used to probe search users' behaviour. Some details of the interface were altered using Google's API, to allow some of the experimental manipulations.

#### 1.4.1 RESEARCH ETHICS

The work presented throughout this thesis has been completed in line with the University of Bath's Department of Computer Science 13-point ethics checklist and to ensure that potential violations did not occur during the collection of data, this was reviewed prior to each study. In Appendix A, an overview of this is provided. While there was a concern that one of the study procedures involved a very minor deception concerning the

---

descriptions of ordering of search results returned was used in Experiment 2 in order to investigate the users' search behaviour, none of the participants objected or showed unease during the post-study debriefing regarding it.

## **1.5 THESIS OUTLINE**

This section presents a summary of the content of each chapter in this thesis, including the studies and contributions where appropriate.

### **Chapter 1 - Introduction**

Chapter 1 has introduced the focus of this research, described the terms that are used throughout the thesis, defined the scope and aims, briefly discussed a number of topics that are explored in the following chapters as well as the methods employed in order to conduct the research.

### **Chapter 2 - Literature Review**

Chapter 2 discusses the existing literature that is pertinent to the investigations carried out in this thesis. The issue about the choice overload effect is pointed out, and the empirical literature on this phenomenon in consumer behaviour and cognitive psychology is reviewed, before accounts of the phenomenon are considered. The chapter moves on to examine search engine users' behaviours and identify how aspects of search engine design influences these behaviours. These two sections are brought together to consider the likelihood of choice overload in search engine use, including a detailed review of the one published experiment to directly address this issue. Then, based on the effects of the ordering of search results, the concerns about a variety requirement and the chance of encountering serendipity in search engine use are discussed.

### **Chapter 3 – Choice Overload in Search Engine Use**

This chapter contributes to the understanding of the effect of a number of search results returned from a search engine on users' satisfaction with their selected options. To this end, the findings from a series of three experiments are presented and discussed. The results investigate the situations where choice overload would affect the users' satisfaction and reveal the determinants of such effect. Furthermore, the role of the



ordering of search returns is investigated in terms of its effect on the users' behaviour and satisfaction as well as how such ordering can mitigate the choice overload effect.

*Experiment 1: Number of Options and Iteration* – The first experiment explores the choice overload effect in an online searching environment with 32 participants, who were asked to perform search tasks to find valuable information for open-ended problems. The effect of the number of options was investigated.

*Experiment 2: Number of Options and Ranking* – In this experiment, 24 participants performed search tasks in a two-factor experimental design in which the number of search returns and the ordering of those returns were manipulated. The results illustrated the effect of the perceived ranking efficiency, which shaped the experimental design in the next experiment.

*Experiment 3: Number of Options, Ranking and Time Pressure* – based on the findings from the previous experiment, in this experiment the 24 participants were performing search tasks while time was used as a constraint and the results revealed that this is an important determinant of choice overload in the search engine use context.

### **Chapter 4 – Variety Seeking and Chance Encountering**

The users' ranking reliance behaviour reported Chapter 3, (i.e., that most of the participants browsed and selected options from near the top of the results list), is the primary motivation for the investigation presented in this chapter. It contributes to the understanding of the effects of a variety of search results in two aspects: users' search behaviour and their chance of encountering serendipitous experience under the conditions where either a single option or three options is/are to be selected.

*Experiment 4: Number of Options and Supplementary web pages* – supplementary web pages are introduced in this experiment as an opportunity to encounter options from different sets of search results. The experiment involved 48 participants who performed the search and selection tasks with sets of six, 24, or 48 options of search returns, with an additional two supplementary web pages being presented for each task in the separated lists. The users' search behaviour and their satisfaction with the selected options were investigated and participants' opinions about the usefulness of the supplementary web pages were explored. The results revealed the requirement of variety was derived from

---

the participants' seeking behaviour. Moreover, qualitative analysis of the perceived usefulness (or otherwise) of the supplementary web pages is reported.

## **Chapter 5 – Serendipity and Literature Search**

This chapter builds upon the main contributions of Chapter 4, with the aim being to explore the benefits of a parallel synonym search facility in the context of chance encountering when the ultimate goal of the search is undefined.

*Study 5: Automated synonym-enhanced search for serendipitous experience* – The study involved asking 12 research students to search for publications related to their research area by using an automated synonym-enhanced search application. After performing search tasks for five consecutive days, the participants were interviewed about their general use of search engines for research purposes and their experience of searching using the novel application. The results from the study contribute to knowledge by revealing the information needs and moderators that affect literature search behaviour as well as demonstrating the usefulness of synonym searches to enhance serendipity encountering.

## **Chapter 6 - Conclusion**

The final chapter provides a review of the main findings and contributions of this thesis. In addition, the limitations are identified and potential directions for future work discussed.



## **CHAPTER 2**

### **LITERATURE REVIEW**

#### **2.1 CHAPTER OVERVIEW**

The primary objective of this thesis is to investigate the effects of the number of search results returned and their ordering on users' behaviour and especially satisfaction, in a search engine use environment. Of particular interest is the extension of the "choice overload" phenomenon to this domain. Choice overload occurs when large choice sets have negative psychological consequences. Users of search engines are often faced with very large choice sets, presented as a list of snippets. The thesis investigates whether this particular design results in choice overload and what factors might moderate such effects.

Thus, the work aims to understand users' perceptions and experiences and to identify factors that affect those in order to elicit how search engine designs can be improved to support users' needs. In order to address such issues, it is necessary to examine the existing literature on user behaviour in relation to search engines as well as the underlying knowledge concerning decision making and satisfaction with decisions deriving from cognitive psychology and consumer behaviour.

This chapter reviews literature and empirical studies relevant to the concerns of this thesis, being organised as follows. In the first part, the choice overload phenomenon is examined, in particular, regarding the positive and negative effects of choices in different contexts including in search engine use. Next, search engine designs are probed in relation to the effect of the number of search results presented and the ordering of

these on user behaviour during a search process and their satisfaction on the search outcomes. Furthermore, this literature review investigates the role of satisfaction as a measure of search outcomes and the factors that affect satisfaction judgement. Subsequently, issues raised from the ordering of search results returned are reviewed and discussed, namely, variety seeking and serendipity. In addition to that examined in this chapter, some literature related to the experimental design is reviewed in the introduction sections of the empirical chapters as the immediate background to the studies reported in them.

---

## 2.2 CHOICE OVERLOAD

In many situations, it has been demonstrated that the ability to choose yields benefits compared to when choice is not available. For instance, in modern health care, a study of placebo painkilling treatments showed that the participants reported less pain when they were able to select between two treatments (Geers et al., 2013). Similarly, in a very different situation, in reading comprehension tasks, it was found that individuals who had a choice between two articles answered questions related to the selected one better than did those who had no options, even though the content of those articles was identical and the only difference between them were the titles (Patall, 2013). In the context of school education, it has been reported that a choice of homework motivated students to do it with a feeling of greater competence and to perform better on a unit test (Patall et al., 2010).

Oppewal and Koelemeijer (2005) also found that when there were more options available this was likely to have positive consequences when compared to the choice being fewer. In their study, the purchase of cut flowers was investigated, whereby a set of flower photographs ranging from five to twelve photographs was sent to a florist's regular customers and each was asked to rate the attractiveness of images they received. Participants were also asked whether they would purchase anything, and if so, which items they would purchase and how many units. The results demonstrated that more options had a positive effect, regardless of the similarity of the items and when the options available contained a preferred item.

Due to the expected positive effects of choice, consumers and policy makers often appear to assume or report that more available options are preferable. However, it has been found that in some circumstances, too many options can lead to negative consequences, such as reduced consumption, deferral or avoidance of choice and lower satisfaction, as a result of which the influential term "The paradox of choice" was coined (Schwartz, 2004). In support of this, recent research has found that choosing from a large number of options may lead to "choice overload", which refers to the negative psychological effects of being asked to do so (Iyengar & Lepper, 2000; Loewenstein, 1999; Iyengar et al., 2006; Carroll et al., 2011; Reibstein et al., 1975).

The choice overload phenomenon was first reported across three experiments by Iyengar and Lepper (2000). The first involved food shopping, where either six or 24 different

jams were presented to buyers in an “upscale retail outlet”. After tasting, each buyer was given a discount coupon that could be used when buying a jam and these were used more by those who tasted from the small set than those tasting from the large set. In their second experiment, students were asked to choose an essay topic from either six or 30 options and then to write a two page essay for additional course credit. The percentage of students who completed their essays and the quality of those essays were both higher for those given the smaller set of options. Iyengar and Lepper’s final experiment required the participants to sample chocolate that was chosen for them or was their own choice and those who could choose were presented with either six or 30 flavours. Having made the choice and eaten the chocolate, they were then offered two options for payment, five dollars in cash or chocolates worth five dollars. Participants who had chosen from only six flavours were more likely to choose chocolates as compensation. Furthermore, those who chose from the large set reported that the selection process was difficult and frustrating. That is, although at first the large number of options seemed attractive and enjoyable to consider, they felt regret and dissatisfaction with their final choice.

It is not immediately clear what is so different about Iyengar and Lepper’s experiments and those of Patail or Oppewal and Koelemeijer. This question will be returned to once some more studies showing choice overload effects have been reviewed.

In the non-financial decision context of volunteering, the effect of a large option set has been investigated (Carroll et al., 2011). In this study, the participants were asked to choose an organisation to volunteer for from either 10 or 30 voluntary organisations, which they were allowed to search for information about before making a decision. The findings showed that those who were presented with 30 options were more likely to defer their decision than those who were offered 10 options.

Park and Jang (2013) conducted an experiment to explore the existence of the choice overload phenomenon for products offered by tourism suppliers. The results revealed that choice overload was likely to occur when the presented products involved more than 22 options, for those participants in this condition were less likely to make a purchase than those not. In addition, when the number of available products in the set was less than 22 items, the participants who made no choice reported more regret than those who did choose. Reutskaja and Hogarth (2009) found that choice satisfaction, both product and process types, are inverted U-shaped functions of the number of options. That is,

---

when their participants were presented with a set of gift boxes containing five, 10, 15, or 30 alternatives, those who chose from the five and 30-option sets reported lower satisfaction with the selection process and the product chosen than those in the other two conditions.

Notwithstanding the findings, as hinted above, the empirical evidence concerning the choice overload effect is somewhat inconsistent, with many studies reporting positive effects of large choice sets and in some cases failing to find any significant evidence for an effect of choice overload. Scheibehenne, Greifeneder and Todd (2010) directly attempted to address this inconsistency in a meta-analysis of 50 published and unpublished experiments in which the number of alternatives was the major independent variable, combining the effect-size of measures such as consumption quantity, unwillingness to choose or satisfaction with final choice. Their results revealed that the mean effect size was close to zero. Furthermore, these authors failed to identify, from the literature, any sufficient conditions for a choice overload effect. In fact, it is a little easier to suggest *necessary* conditions for choice overload, in particular, the non-familiarity of options seems necessary, because otherwise the decision-maker tends to fall back on simple recognition and preference judgements (Scheibehenne et al., 2009; 2010) that make the size of the choice set less salient.

Some studies have directly questioned explicit preferences. Arunachalam et al. (2009) found that when people were asked whether they would voluntarily decrease the options in a set from 24 to six options, a significant portion of the respondents said they would do so, because choosing from the small set would ease the decision process. However, in a follow up study where the participants were offered alternatives between receiving two dollars in cash and choosing one free soda from a set of either six or 24 options, the participants who could select from the large set choose to select a bottle of soda more than those who were offered the small one.

In a study conducted in a wine retail environment, where consumers self-reported their preference for the existing choice of stock or an increase, Zucker & Remaud (2014) found no apparent evidence of the choice overload effect happening. Indeed, they discovered some would rather have more options even though there were already up to 1800 available.



Scheibehenne, Greifeneder and Todd (2009) conducted a series of studies to investigate the specific conditions that may lead to the negative effects of choice overload. In particular, the effect of the number of options, the attractiveness of these, and the effect of having to justify a decision were investigated. The participants in the first study were entered into a lottery simulation and asked to choose options as a prize if they won, that is, between 30 euros in cash or a 40 euro coupon for any restaurant displayed in a list of either five or 30 options. Although the number of participants that selected a coupon for a restaurant was more than those who chose cash, there was no difference in the number of coupons selected from the large and small sets of restaurants. The second study was conducted in the context of selecting charity organisations to donate money to and the results revealed that a large number of options was likely to lead to more donations. The only negative effect found was that an increase in the number of options was likely to increase difficulty in justifying the choice made, especially in terms of defending the selected option against other alternatives. In the final study, the task was to select a classical music album, where participants were allowed to listen to a sample of the first track from their selected album. They were then asked to rate the level of motivation to make a choice, their post-choice satisfaction and post-choice regret. The participants' consideration set (those items they actively considered) and search behaviour were also investigated. Consistent with the former studies, the results revealed no evidence that the participants were negatively affected by choice overload, even though the large set led to an increase in the number of options considered for each task and in the perception of task difficulty.

Three broad explanations have been put forward for choice overload effects, namely, process-product leakage, regret and expectation. These explanations are described here in terms of a single effect that is lower satisfaction with a chosen alternative when the set from which the alternative was chosen was larger rather than smaller. Nevertheless, it is believed that the reasoning can readily be adapted to other dependent measures, such as motivation to consume.

The first explanation is termed here process-product leakage, whereby a large set of items to choose from is likely to make the process of choosing more problematic in several ways. Most obviously, if one assumes that time is constrained, it will mean less consideration must be given to each item (no consideration at all of some items in many cases of search engine use). More subtly, a bigger set of options makes it more likely

---

that some items are hard to contrast (e.g. the second-best is likely to be closer in perceived quality to the best). If the decision process is difficult, the final choice may be viewed as unsatisfactory, because the process that led to it is unsatisfactory in some way. In addition, it was suggested that even if an experimental participant is responding to a question about their satisfaction with a consumed item, it seems plausible that their response may be less specific than the question's wording (Scheibehenne et al., 2009) in that the satisfaction could be influenced by the perception of the decision process. The process-product leakage explanation predicts that decision time will be a moderator of choice overload effects, as has been proposed by Haynes (2009). In his research, participants were asked to choose a prize to be entered in a draw from a set of prize descriptions; either three or ten were presented. The results showed that under the time pressure they felt the decision was difficult and reported less satisfaction, especially with a large set of options. The second, related, explanation for choice overload is regret, i.e. with regards to not opting for the unchosen items (Schwartz, 2004). The argument is that the more items rejected, the more likely that the decision-maker regrets doing without some of those items, i.e. a counterfactual, if-only, response negatively impacts on the post-hoc evaluation of the chosen item. Finally, the explanation of the effect of choice-set size may be on expectations. That is, if a choice set is large, the decision maker may expect a better outcome than were it small, and this expectation works to set a standard against which post-choice comparisons are made. This explanation is in keeping with evidence that suggests participants have an a priori (pre-choice) more favourable reaction to large sets of alternatives (e.g. Iyengar & Lepper, 2000).

Intuitively, all three of these psychological processes seem plausible in the case of choosing information sources from the set of query returns by a search engine. On this basis, it seems reasonable to predict that choice overload is an important issue for search engine design. However, it appears that the effect of choice overload is somehow patchy across the investigations with a wide range of moderators and covering various choice contexts. This thesis is interested in the use of search engines, where choice sets are very large and where user satisfaction with found items is an important measure of interaction quality. The general literature on choice overload is not very conclusive concerning the question: Will users of search engines suffer from choice overload? To further frame the question, and the parameters that may affect it, we will now review literature on behavioural and decision-making aspects of search engine use.

### 2.3 SEARCH ENGINE

#### 2.3.1 SEARCH ENGINE USE

Information seeking is an iterative process, in which understanding of a situation can develop during a search through interpreting and evaluating information, developing understanding, forming new questions, and seeking new information to answer the questions (Belkin et al., 1995; Marchionini, 1997; Blandford & Attfield, 2010). This interaction with information is described by Blandford and Attfield (2010) as an “information journey”, whereby people iteratively recognise a need for information, find information, assess that information based on their goals, and use the interpretation derived from that information to support their decision in performing further actions. During an information seeking process, searchers’ interactive activities might vary based on the context and their goals. For example, in professional disciplines such as healthcare, the need might be to gain information about the current best evidence of diagnoses and treatment plans, whereas in a tourism context, individuals might seek for variety of information about tourist attractions from different sources. Moreover, levels of expertise in both the subject area and searching skill could influence strategies used in information seeking process (Hsieh-Yee, 1993; Fields et al., 2004).

Web search engines are used for various information seeking purposes, ranging from very simple daily life enquiries, such as finding a pancake recipe to much more complicated search tasks, for instance, exploring solutions to problems in academic research. The underlying intentions of web searches have been classified into three basic types by Broder (2002). The first is *navigational*, which refers to the goal of reaching a particular web site that is known or unknown by a searcher. Secondly, there is the *informational* search, the objective of which is to find information that is assumed to be provided by one or more web sites. The final type, *transactional*, is where a searcher has the intention of performing further actions, for example, browsing songs for downloading, finding a sharing server, or shopping. Further investigation on search users’ intent was conducted by analysing search query logs (Rose & Levinson, 2004) and the results revealed that users were more likely to search for information rather than navigate web sites or find resources for transactional activities. Of course, the relative frequency of types of search do not say much about their relative importance, even if they do go some way to characterising search behaviour in general terms. In Rieger’s

---

study (2009), the use of search engines among students and academic staff were explored. This study employed three data-gathering methods: a web-based survey, interviews, and search log reviews. The findings showed that search engines were used in various search contexts including academic research, health, and hobbies. The participants were generally satisfied with search engines and had trust in the ranking algorithm, such that the information they needed would appear within the first few search results pages. In addition, it was found that the types of search task had a significant effect on the types of query reformulation (Liu et al., 2010). In this web-based information search experiment, users' queries and their reformulation types as these related to task types were investigated. The results revealed that the users tended to use specific queries for a simple search task that required a single piece of information. On the other hand, for an information gathering task, word substitutions and synonymous reformulations were more frequently performed in order to find multiple aspects of a single topic or multiple topics.

### **2.3.2 SEARCH ENGINE DESIGN**

In most cases, search engine users are presented with such extensive lists of search returns that it would seem impossible for them to assess all available results, which in turn appears to affect their search behaviour. As reported by a comparison of nine search engines' transaction logs, Jansen and Spink (2006) found that the proportion of US-based Web search engine users who viewed only the first SERP (Search Engine Returns Page) was 73% in 2002. Although this may be an indication of the high efficiency of search engines in that the ranking can represent the users' needs to a satisfactory extent, it could also be the case that it is impossible for users to evaluate a large number of options, because their processing capacity is somehow limited. Some users may try to reduce cognitive load by adapting search strategies that decrease the number of search results returned (Xie & Joo, 2010; Jansen & Spink, 2006; Spink et al., 2001; Kim & Allen, 2002). In addition, the users' domain knowledge and experiences on web search have been found to be moderators that influence their use of search strategies (Hölscher & Strube, 2000; Thatcher, 2008; White et al., 2009; Cole et al., 2011; White & Morris, 2007).

It was found that users change strategy when seeking information based on the number of results returned from a search (Stelmaszewska & Blandford, 2002). This particular

study monitored users' search behaviour when working with digital libraries. The findings revealed three kinds of search returns that affect users' further search actions. First, the response for a "no matches" result was that users preferred to change search terms and would change to use a different library if their attempt failed after a few times. Second, in the cases of "too many" results, the study reported that some users viewed through the returned items up to three SERPs at most, whereas others appeared to be overwhelmed and tried to reformulate their search terms. Additionally, in these cases the users appeared to question their abilities to devise effective search terms. Finally, with an "OK number" or manageable number of results users would scan them and make a relevance judgement, which was followed by examination of the selected items in detail based on this judgement.

Regarding a large amount of results returned due to the great volume of data sources, some features of search engines that have been designed for managing such returned sets might mitigate the effect of number of options to some extent. First, in many the long list of returns is broken into separate SERPs, so that the length of the list is arguably less salient. Studies have shown that paginated lists lead to better performance and memory than uninterrupted, scrollable ones (Schwarz et al., 1983; Bernard et al., 2002; Piolat et al., 1997). In particular, studies have suggested that people are able to locate relevant information and remember details better when browsing through pages than when scrolling down a single page. On the other hand, the advantage from the pagination could be explained by consideration set theory, whereby in order to simplify a decision making process, consumers mostly use some relatively quick method to reduce all available options to a smaller group called a consideration set. Options in this set are those that decision makers consider before selecting a final choice (Shocker et al., 1991; Narayana & Markin, 1975). Consequently, the pagination features of search engines may be perceived as an effective method to reduce the options from a long list, which make the search process less difficult and more satisfactory.

Further, the fact that search results are ordered may well affect user behaviour and search satisfaction. Users may be very confident that reliable ordering will lead them to the most relevant options and consequently, they may be unaware of the length of the returned list, only selecting items that presented early on. This is supported by evidence from a number of studies, which found the ordering feature of search results presentation affects judgements of option relevance (Eisenberg & Barry, 1988; Keane et al., 2008;

---

Pan et al., 2007; Lorigo et al., 2008; Guan & Cutrell, 2007). The influence of presentation order was examined by Purgailis and Johnson (1990) through experiments, wherein students were given essay coursework and instructed to use the provided system to search for relevant citations. The list of search results was randomly generated based on the keywords that were entered and they were asked to examine all the documents that were listed in SERPs to form judgements regarding their relevance. The results revealed that when the list contained more than 15 items, documents presented early on the list were likely to be considered as more relevant. However, such a presentation effect did not occur when there were less than 15 options in the lists. The findings were later on supported by Huang and Wang (2004), who conducted an experiment for a similar context. Again, the results showed that the order of options significantly affected the relevance judgement when only 15 to 30 options were presented.

Although it appears that the ordering of search returns provides the advantage to users that it simplifies the search process in some way, the fact that most trust the ordering of search results and judge web page relevance based on position has raised concerns regarding the ranking's reliability. Consequently, researchers have investigated the reliability of search engine ranking algorithms with the aim of improving their efficiency in various respects. For example, based on the concern that the orderings may be unreliable relative to users' needs, they have been investigating how user behaviour can be incorporated to improve web search ranking (Agichtein et al., 2006; Bar - Ilan et al., 2007; Gao et al., 2009) and how a search results list can be diversified to maximise the coverage of different user intentions (Zheng et al., 2012; Santos et al., 2010, Drosou & Pitoura, 2010, Dou et al., 2011). Furthermore, in the context of chance encountering, researchers have questioned whether the reliance on the orderings of some particular search engines may limit the users' opportunity to discover valuable options that are treated as low quality by some ranking algorithms (Goldman, 2006; Pandey et al., 2005; Cho et al., 2005; Toms & McCay-Peet, 2009).

The work presented in this thesis focuses on the interactions between users and search engines. In particular, it tackles the issues about the effects of a large number of search results returned and the effect of ranking algorithms or the ordering of the returns, namely choice overload, variety seeking and serendipity. Moreover, how system designs can be improved to provide better support for the users' needs regarding these issues is explored. Instead of trying to assess the efficiency of search engines by measuring recall,

the coverage percentage of relevant results, and precision, the level of relevance (Raghavan et al, 1989; Brin & Page, 1998; Gordon & Pathak, 1999), the work reported in the following chapters concentrates on users' satisfaction as the evaluation of their search engine usage. In the next section, the different dimensions of satisfaction are examined along with the factors that may affect such judgements in the search engine use context.

### **2.3.3 SEARCH BEHAVIOUR MODELS**

The issue, that search returns are ranked or ordered, and that the order is likely to be relevant to the user's task is a crucial issue for the studies in this thesis. Some cognitive science research has investigated models of the choice process that help to illustrate the vital effects of ranking on decision making in this context.

In particular interactive search processes have been investigated with the aim of modelling search behaviour and strategy (Brumby & Howes, 2003, 2008; Cox & Young, 2004; Fu & Pirolli, 2007). Process models of this kind are not the major concern of this thesis, so their mechanisms will not be reviewed in detail, but they allow some insight into the behaviour of search engine users that is relevant to the data reported, so this review will attempt to extract a general behavioural framework.

The earliest relevant model developed by Cox and Young (2004). The task modelled by these researchers is to choose the one correct item from an ordered list of items in a menu (c.f. the returns of a search engine or a menu of links from a web page). This is clearly not identical to the typical choice situation in choice overload experiments, in that in those experiments there is no single correct choice, but both situations involve a user trying to decide which of a list of items to select.

Cox and Young's (2004) model assumed that items in the list are assessed in sequence and that after the current item was assessed based on its relevance to the search goal and contrasted with estimated relevance of previously assessed items, participants would then perform one of the three actions: select the current item because the benefits of doing so are judged to outweigh the costs of any further assessments, assess the current item in more detail, or move on to assess another item. This model allows a variety of behaviour to emerge, depending on the judged quality of items being assessed. Especially, the model can explain how a very good early item may lead to termination of

---

search without further consideration of items. Also, the model can explain a pattern of search whereby users inspect the next few items below a good item before returning to and selecting this item.

Brumby & Howes (2008) conducted two experiments in order to explore the option-assessment process and the selection behaviour during goal-directed search by employing eye movement tracking. Their first experiment investigated the effect of the degrees of relevance of options in the sets on participants' search strategies, and in that participants were asked to search for a target item that was relevant to a given goal statement from the list of 16 options for each task. In addition, they were informed that there was only one correct item in the set and the rest of options were distractors. The results from this experiment revealed that when the distractors were less relevant to the task goal, participants gazed at fewer items and often skipped over options after they found a candidate item. However, when the distractors were more relevant, more items were assessed and reconsidered before making a selection. In their second experiment, the effect of the location of target item on search behaviour was examined. The results showed that when the target item was presented towards the top of the list, participants tended to assess more items located further down, often in the skipping items manner, before selecting it. In addition, participants were more likely to select a target item when first encountered if it was positioned towards the bottom of the list. From these experiments, it was demonstrated that searchers did not typically consider all of the available items or stop assessing items when the value of most recent one was above a threshold. The latter behaviour is inconsistent with the satisficing heuristic proposed by Simon (1995), which suggested that a search process would continue only until the defined threshold was met. However, it is consistent with the general idea of bounded rationality and with one property of the satisficing process, namely that the best item may not be found.

Fu and Pirolli (2007) built a computer simulation that shares many properties, with the models explored by Cox and Young (2004) and Brumby and Howes (2008). SNIF-ACT; Scent-based Navigation and Information Foraging in the Adaptive Control of Thought cognitive architecture (Fu & Pirolli, 2007) is implemented on the ACT-R; Adaptive Control of Thought–Rational architecture (Anderson et al., 1997) and uses spreading activation among a word network to model the assessment of relevance of each link in a list of possibilities. In the SNIF-ACT 2.0 model, such assessments are



made relatively and in sequence (i.e. are influenced by the assessment of earlier items in the list through a Bayesian updating mechanism), such that the model might assess several items in a list, but not all, and might thereafter return to the “best so far” which is assumed to be perfectly remembered. SNIF-ACT can thus exhibit the range of behaviours modelled by Cox and Young (2004) and reported by Brumby and Howes (2008).

These behavioural possibilities seem, intuitively, quite likely in the case of using a search engine to find a useful web page. In most situations, search engine users are less likely to browse through all of the options returned, especially when the set is extremely large. Therefore, it was anticipated that users would perform their search as described by previous work to some extent. Firstly, they may stop searching once the value of the most recent option has reached a subjectively defined threshold. Secondly, they may continue browsing more options, even though the good enough option is found, in order to increase confidence in their decision and reduce the risk of neglecting any potential high-value options. Above all, ordering of options may affect a relevance assessment of each option in this context such that users may be biased on estimating relevance of options ranked higher in the list, which lead them to consider only a small proportion of options that are located towards the top.

### **2.4 SATISFACTION**

People often come across situations that force them to choose and the perception of a decision success or selecting the “best” choice varies from one individual to another. Satisfaction is one of the measures that represents the overall judgement with the decision made and in the case when this is technology-assisted, satisfaction also represents the effectiveness of the assisted system (Gatian, 1994). The present thesis focuses on the role of satisfaction as a measure for evaluating choice decisions in a search process so as to investigate the effect of choice overload. The following subsections describe the role of satisfaction in measuring the outcome of choice decisions. To begin with, the literature that discusses definitions and different dimensions of satisfaction are examined.

---

#### **2.4.1 SATISFACTION DEFINITIONS AND DIMENSIONS**

In the situation that a decision is complex and the “right” one may be defined differently among individuals, the level of satisfaction on a decision outcome has been proposed as being a valid summary measure for a decision evaluation. Oliver (1996) defined satisfaction as an experiential judgement of outcomes as a sense of fulfilment and suggested that the value of a decision outcome, such as, quality of product, enjoyment, fulfilment, fun, and security, be compared in order to enhance satisfaction. Based on the existing literature and consumer interviews, three general components of satisfaction can be identified: response, focus and timing (Giese & Cote, 2000). First, consumer satisfaction is an emotional or cognitive response, which varies in intensity depending on the situation. Second, the response of consumer satisfaction relates to a particular focus, such as expectation, product, consumption experience or purchasing decision, whereby these are compared with some specific or general standard. The last component is timing of the response, such that satisfaction could change during the time of measurement. For example, satisfaction of the selected product measured before and after it was experienced could be different. Based on these components, it could be inferred that satisfaction judgement involves comparing the value or the particular focus of the outcome to a set of goals or standards, and that it could be affected by various factors depending on the decision situation.

Recent research has focused on evaluating a decision by measuring individual satisfaction judgement and in general, much of this work has been concerned with post-decision satisfaction. In a study by Sainfort and Booske (2000), participants were asked to access information about alternative health plans via a computer-based system. More specifically, they were tasked with entering their plan choices based on the provided information, and then to rate their attitudes towards their decisions through a decision attitude scale. This post-decision satisfaction was measured right after they had made their decision with the outcomes having not yet been experienced. The analysis of the decision attitude scale revealed that decision satisfaction comprised three underlying dimensions: satisfaction with choice, usability of information, and adequacy of information, which hence, meant that it is multidimensional. Additionally, the researchers suggested that other potential dimensions of post-decision satisfaction may be identified in a different decision context. For example, in health-related decisions the dimensions of satisfaction might be classified in terms of decision responsibility, e.g. a

doctor or a patient will be responsible for the outcome of the selected treatment as well as self-efficiency in terms of a successful decision implementation, i.e. satisfaction in the efficacy of employing the selected treatment. From the study of Sainfort and Booske (2000), it emerged that post-decision satisfaction is not purely based on the decision outcome, but also influenced by other factors involved in the decision process, such as the usability and suitability of provided information and the domain knowledge of the decision maker. In order to measure satisfaction, it is necessary to identify dimensions by which it will be judged in order to provide more robust interpretation of satisfaction results and the factors that influence them.

From the literature review so far, it appears that satisfaction judgement can be categorised into two main dimensions: process satisfaction and product satisfaction. The former, process satisfaction, refers that satisfaction is influenced by the factors in the decision process, such as perceptions of available option sets, domain knowledge of the decision makers regarding the context, decision strategy, enjoyment during making a decision, etc. On the other hand, product satisfaction, is judged based on the attributes of the selected product or service, such as its quality, usability, attractiveness, suitability, etc. These underlying two dimensions are our main focus when examining the determinants of satisfaction.

Prior research has demonstrated that factors involved in the decision process indirectly influence the judgement of satisfaction through the individual goals of: maximising the ease of decision justification, increasing confidence, minimising the experience of negative affect, minimising decision evaluation costs, and minimising anticipated regret (Heitmann et al., 2007). In these authors' study, consumer surveys on purchasing decisions and search behaviour were conducted with people who had purchased an electronics product within the past three months. Satisfaction judgement was measured in the two focal dimensions, namely, process satisfaction and product satisfaction. Five decision-process factors that were anticipated to influence the aforementioned choice goals were classified into three groups, including individual factors (predisposition towards maximising and product category knowledge), choice-set factors (product overload and product complexity), and a social factor (ambiguous social reaction). The findings demonstrated that process satisfaction was primarily driven by anticipated regret and final negative effect, whereas the product satisfaction depended on choice confidence, evaluation cost, and that regarding the decision process. In terms of the

---

effects of the choice-set factors, product overload was found to be correlated with anticipated regret, evaluation cost, and justifiability, whilst product complexity correlated with choice confidence, justifiability, and final negative effect.

It would appear that decision-making goals, which are affected by decision-process factors, can motivate the judgement of satisfaction. For example, the goal of minimising anticipated regret may be negatively affected by the overload of products in that enlargement in the number of products available, could increase the regret about being unable to consume the unchosen ones. As a consequence, this could result in less satisfaction with the selected option (Heitmann et al., 2007). Moreover, the satisfaction on the decision outcome is affected by the satisfaction regarding the decision process. As noted by Botti (2004), differences in product satisfaction may emerge during the decision-making process before the decision outcome is experienced and the factors involved in such a process can be the moderators of the product satisfaction. In the next subsection, the aim is to examine the decision factors and their effect, directly or indirectly, on process and product satisfaction.

#### **2.4.2 DECISION FACTORS AND THEIR EFFECTS ON SATISFACTION**

Regardless of the quality of the decision outcome, many factors can influence the evaluation of the decision satisfaction (Churchill & Surprenant, 1982; Oliver & DeSarbo, 1988; Tse & Wilton, 1988). This section describes decision factors and their effects on satisfaction judgement that have been investigated in different choice contexts. The factors reported below cover the characteristics of the choice set, individual differences among decision makers, and time pressure in a decision process.

##### **2.4.2.1 NUMBER OF OPTIONS**

Making a decision from a set that has a lot of options available generally requires more time and effort, given that all available need to be considered. Thorough consideration may result in a high quality outcome and hence, lead to satisfaction. However, research has demonstrated that too much choice can negatively affect satisfaction with the selected option. For instance, Messner and Wänke (2011) found that the number of available options is associated with frustration during decision making and post-decision regret. The results from their experiment revealed that choice satisfaction decreases when consumers make their decision from a large set of options. However, when

participants were distracted during the decision task, they were more satisfied with their selected option from the large set (24 options). It seems that the effect of the number of options is also moderated by other factors, such as time pressure (Haynes, 2009; Oulasvirta, Hukkinen & Schwartz, 2009), maximising tendency (Iyengar, Wells & Schwartz, 2006), product similarity (Zhang & Fitzsimons, 1999), and perceived information overload (Messner and Wänke, 2011). The effects of a number of options have also been examined in association with the following decision factors.

### **2.4.2.2 ATTRIBUTE SIMILARITY**

A series of experiments was conducted to investigate the effect of attribute similarity between options and the effect of option limitation on process satisfaction (Zhang & Fitzsimons, 1999). The findings demonstrated that choosing from a set of options with corresponding attributes yields more satisfaction compared to selecting from an option set containing different, unique, ones. Furthermore, limitation of options decreases the process satisfaction when these share corresponding attributes. However, they found that option limitation has no effect on satisfaction regarding the decision process when each has unique features. The explanation for these outcomes put forward is that a set of options with a high level of similarity increases the ease of comparison, which expands the ability to process attribute information and hence, yields satisfaction with the decision process. Further investigation by these authors elicited that formation of a consideration set, i.e. selecting three out of five options, increases the process satisfaction when choosing among unique-attribute ones. In addition, they found if the limitation of options in the initial set was reduced owing to lack of availability, this decreased the satisfaction regarding the selection process.

Thus, it would appear that although similar attributes between options facilitates the ease of comparison and the ability to process more options, the relationship between the ease of comparison and the amount of available options can be complex. For example, in the decision context that a set of options is extensive the level of difficulty in comparison may increase, which may consequently affect the process satisfaction, especially when decision makers are required to justify their choice (Scheibehenne et al., 2009). Other evidence supporting the positive effect of attribute similarity was found in the context of smart phone consumption. Wang & Shukla (2013) conducted a survey study to investigate the effects of choice overload, perceived attribute similarity and information ambiguity on choice confidence, evaluation cost, and final negative effect, i.e. nervous,

---

anxious, and jittery. They found that ambiguous information and choice overload reduced choice confidence, whereas perceived attribute similarity between products and brands increased it. Thus, the evidence suggests that attribute similarity of options may increase the satisfaction regarding the decision process through the increase in choice confidence.

#### **2.4.2.3 DECISION CONTEXT FAMILIARITY**

Familiarity of decision context may indirectly affect the decision satisfaction in both the process and product dimensions. In particular, when making a decision in an unfamiliar context, consumers may have less strong preference on the outcome. The results from a service consumption study (Söderlund, 2002) demonstrated that the level of decision-context familiarity can determine the level of the product satisfaction based on service performance. That is, it was found that when the perceived quality of the performance is high, consumers who are familiar with the decision context are more satisfied with the outcome than those who are unfamiliar. Conversely, when the perceived quality of the service performance is low familiarity lead to less satisfaction with the product satisfaction. Further, regarding the effect of familiarity, Mogilner et al. (2008) discovered that the satisfaction of consumers who are unfamiliar with the decision context can be positively influenced by the presentation of the options regardless of the content or quality of the selected options.

#### **2.4.2.4 EFFORT**

The perception of effort that a decision maker spends in a decision-making process has an effect on the decision satisfaction (Cardozo, 1965). For, when the decision task requires a high level of effort, decision makers feel less pleasant, more exhausted, and more frustrated, which leads to less satisfaction with the decision process. However, the satisfaction with the product may be higher when considerable effort is perceived to have been exerted in obtaining the product than when this is modest. In a study by Bechwati and Xia (2003), a simulation of an online job search was conducted to measure the satisfaction with the job search process and the perceived effort saved by using an electronic aid. The findings showed that the perception of effort saved in the search process was positively associated with the process satisfaction.

### **2.4.2.5 MAXIMISING TENDENCY**

Individual difference in the tendency to maximise a decision outcome seems to be one of the determinants that decreases decision satisfaction. That is, despite the effort required to achieve the goal, a decision maker with high maximising tendency is likely to spend extra effort in a decision process. Maximising tendency may lead to a better outcome in terms of the extent to which more effort is spent (Cardozo, 1965). However, thorough consideration of this form in a decision process can result in outcomes that are judged unsatisfactory, because they fail to meet expectations. The findings from an investigation into the job searching behaviour of final year students (Iyengar et al., 2006) showed that those who applied for more jobs and gathered more information, especially from external sources, were less satisfied with the job they were offered even though they could get a higher salary than those who exerted less effort. Typically, consumers with greater maximising tendency tend to expand the number of possibilities by considering more options, which increases individual expectation and thus, the chance of encountering regret. Perhaps unsurprisingly, therefore, a maximising tendency has been empirically found to be related to regret and negative satisfaction (Schwartz et al., 2002). In sum, consumers who are likely to maximise their decision outcomes are more regretful and less satisfied with the selected options than those who make a decision when the options are acceptable.

### **2.4.2.6 TIME PRESSURE**

A negative effect of time pressure on the product satisfaction has been reported (Haynes, 2009). This author demonstrated that selecting an option from a large set under time constraints leads to less satisfaction with the selected option. However, despite the finding that time pressure increases decision difficulty and frustration, it appears to have no effect on enjoyment in a decision process. As noted by Dhar and Nowlis (1999), time pressure increases difficulty in a decision-making process and influences decision makers to use non-compensatory decision approaches in the selection decision. This may be the reason that decision quality decreases under time pressure and hence, leads to less satisfaction with the choice made. Hahn, Lawson and Lee (1992) measured the decision quality when the information load was increased and the results revealed that without time constraints, the quality of the choice decision gradually increased as more information was presented. Conversely, when the participants were making their choice decision under time pressure, the quality declined relative to the increase in the

---

information load. Recent research has illustrated that perceived time pressure influences the strategies used in information seeking and negatively affects satisfaction with the decision process (Crescenzi et al., 2013). In the study, the participants were asked to search for and bookmark web pages, with the search task having to be completed within fifteen minutes. The findings showed that those who perceived high time pressure reported the perception of high task difficulty and less search process satisfaction than those who did not.

#### **2.4.3 SEARCH SATISFACTION IN A COMPUTER-BASED ENVIRONMENT**

In a situation that computer-based information systems are used to support decision-making processes, decision satisfaction may be affected by the perceived efficacy of the system. Gatian (1994) conducted a study examining the effect of a system on the users' satisfaction, efficiency and decision-making performance. In the study, the use of a financial accounting and accounts payable application system in 39 universities and colleges was investigated. The results revealed that users' decision performance was improved by increasing the effectiveness of the system in that they could make better decisions when more relevant information was available. Furthermore, the findings showed that users' performance and efficiency were strongly positively correlated with their satisfaction, which implies that improving the effectiveness of the system can also increase their decision satisfaction.

The moderators of user satisfaction were examined by Bharati & Chaudhury (2004) when they investigated decision satisfaction in web-based decision support systems. The participants were required to perform a decision task on a web site and rated their satisfaction with the decision made, system quality, information quality and information presentation. Web sites with different web-based decision support systems were randomly assigned to them and the findings revealed that information quality, i.e. accuracy, completeness, relevance, information content needs, and timeliness, was directly and positively correlated with outcome satisfaction. It also emerged that the system quality, i.e. ease of use, convenience of access, system reliability and flexibility, directly affects the decision satisfaction. However, a positive change in information presentation, i.e. graphics, colour, presentation style, and navigational efficiency, does not increase the outcome satisfaction.



It is worth noting that in previous studies the effect of information systems on the satisfaction regarding the decision process was not examined. Despite the quality of information provided by the system, presentation style may affect the decision process satisfaction especially when the amount of information is extensive. Further investigation is, therefore, needed in order to provide better support for the system users. Research has demonstrated that decision satisfaction can be used to evaluate the decision process, decision outcome, and decision support system efficiency. The main focus of this thesis is to investigate satisfaction judgements of search outcomes in order to identify the factors that influence user behaviour and satisfaction in search engine use. It is this researcher's opinion that these issues once unpacked should be brought to the attention of search engine designers as they could help them to improve the efficacy of their systems.

### **2.5 CHOICE OVERLOAD FOR SEARCH ENGINES**

In terms of post-search judgements, providing a large number of search returns would appear to have negative effects on users' satisfaction and confidence in the selected options, which, as mentioned above, is referred to as "choice overload" or the "paradox of choice" (Iyengar & Lepper, 2000; Schwartz, 2004; Oulasvirta et al., 2009). Regarding information seeking using a search engine, it seems clear that these necessary conditions for a choice overload effect are met, at least in many situations especially novel searches, the returned choices will be unfamiliar to the searcher, and the choice set is often large by any standards. However, some typical search engine design features may work against choice overload. First, as mentioned above, the fact that the choice set is ordered (however unreliably relative to the searcher's needs) is a factor that could well affect the decision maker's response. That is, if users are very confident that the ordering is a reliable guide to value, then the length of the list may seem irrelevant, and the tendency to select items from early in the list may mean that the list of options appears *de facto* small and manageable. This argument is supported by evidence that Google users typically select from very early in the list of returned pages (Pan et al., 2007) and further, by studies that show that the judged relevance of documents decreases down lists of more than fifteen documents, even if they are randomly ordered (Huang & Wang, 2004; Purgalis et al., 1990). Moreover, based on the models of search behaviour and strategy (Brumby & Howes, 2003, 2008; Cox & Young, 2004; Fu & Pirolli, 2007), search engine

---

users may stop searching once a very good early item is found without further considering more items and be satisfied with their search outcomes.

Further, in many designs, the list of returns is broken into separate pages, such that its length is arguably less salient, for studies have shown that paginated lists lead to better performance and memory than uninterrupted, scrollable ones (Bernard et al., 2002). It is concluded on the basis of the existing empirical data that it is not possible to predict, with any confidence, whether search engines will produce choice overload. For further guidance, it is necessary to look at the single published study which models a search engine scenario (Oulasvirta et al., 2009).

Oulasvirta, Hukkinen and Schwartz (2009) conducted an experiment to investigate this prediction and reported the negative effect of too many search results being returned. Their study used paper-based materials to model aspects of a typical search engine scenario and because aspects of their study were an important guide to the design of the current work, they are reported here in some detail.

Participants were given three kinds of “realistic” search tasks: Simple facts, such as “Find out which country is located at the highest altitude”; Problems, with open-ended answers, such as “What determines the cost of railway tickets in Europe?”; and Preferences, such as “Find your favourite novelist’s homepage”. For each search task a participant was provided with a printed page containing a query for that task and a set of results. Half of these pages were taken from Google, using its formatting conventions, whereas the other half used an invented search engine, with different terminology and layout but the same content of search results. The main independent variable was the number of snippets returned, either 6 or 24. Independent variables were manipulated in a within-subject design. The participants had to select a single snippet from the returned set, within 30 seconds, and then (without consulting the actual web page) rate their satisfaction with the choice as well as their confidence that they had made the “correct” one. There were no significant differences between the search engines, but a significant choice-overload effect was found, with participants reporting greater satisfaction and confidence when they chose from only 6 snippets. Post hoc analysis suggested that this effect was limited to the simple fact and problem task types.

However, several features of Oulasvirta et al.’s study limit the generalisation of the conclusions to real search engine use. Firstly, paper presentation is obviously an

approximate model in several ways, for example, it does not allow users to specify their own queries or to view any of the found web pages, both of which seem likely to be critical determinants of satisfaction in real scenarios. Secondly, limiting the judgements to snippets rather than the linked-to pages seems to make process-product leakage a very salient determinant of satisfaction. That is, when no end product is actually experienced, what else can a satisfaction judgement rely on except process, or else the snippet information itself, the very basis of the choice. Furthermore, the time limit of 30 seconds per task is quite severe, for it amounts to 1.25 seconds per item in the large set condition, which would also have been likely to engender dissatisfaction with the process. This account is supported by participants reporting that when they chose from the six-option list they thought more carefully about their decision.

Further, in such an ordered environment as a search engine, it is generally accepted that people mostly inspect options in the order they are presented. Thus, a deeper search through a long list of search results returned implies that lower quality options being considered, which consequently could have a negative impact on satisfaction. As noted by Diehl (2005), searching too much in an ordered environment could lower choice quality. A large number of options on offer has been suggested as one of the factors that triggers users to increase the number of considered options. The resulting poor quality under these circumstances, as explained above, could account for the choice overload found by Oulasvirta et al. (2009) and the accompanying lower satisfaction with the selected options.

As mentioned above, there are a number of factors that could affect users' satisfaction with their chosen options. Hence, further investigation on the search satisfaction is required in order to explore whether the number of search results returns is one of those factors that lead to a negative consequence in search engine use.

### **2.6 VARIETY SEEKING**

In this section, literature related to variety seeking is examined in order to provide the background for the work presented in Chapter 4, which focuses on exploring this behaviour by search engine users. The section begins with investigating the motivations of the desire for variety and then discusses the effect of variety seeking on the satisfaction with the decision outcome.

---

Variety seeking behaviour has been considered mostly as a consumer motive. In an interdisciplinary review by McAlister and Pessemier (1982), this was divided into two main variations: derived and direct. Derived varied behaviour refer to behaviour that is influenced by external or internal forces that are unrelated to individual preference, such as *multiple users*; e.g. members of a family preferring different objects, *multiple situations*; e.g. different locations of consumption, time constraints, the quantity consumed, *multiple uses*; e.g. different types of soups used as the basic ingredients for different dishes. On the other hand, direct varied behaviour pertains to behaviour that is influenced by external and internal forces that are directly motivated by individuals. This can be divided into interpersonal motives (external) and intrapersonal motives (internal). *Interpersonal motives* refer to needs for personal identity and group affiliation in society, while *Intrapersonal motives* cover individual stimulation, such as exploration and novelty seeking. Raju (1980) suggested that different levels of individual stimulation can be defined by three factors: the desire for the unfamiliar, the desire for alternation among familiar alternatives, and the desire for information. The author suggested that the different level of individual stimulation is the moderator that affects variety seeking behaviour. Kahn (1995) reviewed the key findings about variety seeking behaviour among consumers and identified three basic motivations for this: internal stimulation, changes in the external environment, and uncertainty regarding future preferences. Firstly, variety seeking may be motivated by an internal stimulation, such as satiation from consumption experience, which may force consumers to react based on their preferences. For example, they may be satisfied with the current selected options, but may also want to try new alternatives from among familiar sets or different sets so as to satisfy a desire for novelty. Secondly, when the external environment changes consumers may respond to this by seeking alternatives, for example, they may need to look for other products when their preferred-product is somehow unavailable. Finally, the uncertainty about what future preferences will be may influence consumers to hold as many different options as possible.

The effect of such uncertainty on variety seeking behaviour was reported by Simonson (1990), who investigated consumers' purchases in two situations: simultaneous and sequential purchases. That is, the participants in the sequential choices condition were told to assume that they were doing their daily food shopping for consumption within a day. Under this condition, the participants were asked to choose one item per category

from a total of eight categories and the process was repeated for three times. On the other hand, in the simultaneous choices condition, the participants were told to assume they were doing their shopping for the next three days and had to choose three items from each category. The results revealed that those who made choices sequentially were less likely to select variety than the subjects who simultaneously made multiple choices. In addition, it was indicated from the study results that selection of variety was used as a simplifying heuristic by consumers who had difficulty determining their preferred items. This study demonstrated that consumers who made simultaneous purchases were uncertain about their future preferences and tended to choose a variety of preferred items rather than multiple replicates of the most preferred ones.

When the desire for variety has been fulfilled, consumers seem to be more satisfied with their choice than otherwise. Moreover, as researchers have demonstrated, consumers' perception of variety among products increases their consumption quantity and enjoyment of the process (Rolls et al., 1981; Kahn & Wansink, 2004). This may be because they are more likely to find what they want among a varied set of options. Thus, it would appear that a larger set of options will give the perception of a greater variety and hence, increase the probability that those available will somehow match with the criteria sought. However, research has shown that the perception of variety is less likely to depend on the number of options than on prior experiences and expectations. As noted by Broniarczyk, Hoyer and McAlister (1998), a twenty-five per cent reduction in the number of options in grocery stores did not yield a negative effect on the perception of variety. Moreover, this reduction may be increased up to fifty per cent without such negative consequence, if favourite items are available in the reduced set and the categories of options are held constant. Further investigation was undertaken by Mogilner, Rudnick and Iyengar (2008), who found that decision satisfaction was influenced by the presence of categories regardless of their content or the number of available options in each. In addition, a larger number of categories was perceived as a representation of a higher degree of variety among the available options. Moreover, decision makers who were less familiar with the options domain tended to rely more on the category-based presentation and perceived this as demonstrating that more variety was available.

Although the perceptions of options variety may increase consumer satisfaction with the decision process and outcome, an extensive level of this can have a negative

---

consequence, whereby they become confused or frustrated in the choice process (Kahn et al., 2013). However, when a small degree of variety is presented, the attractiveness of the options set may be limited. Therefore, identifying the optimal degree of variety is essential in order to manage consumer satisfaction. Chernev (2006) suggested that the required level of variety might be based on individual sub-stages of the decision process that consumers are engaged in. More specifically, the author found that if consumers viewed the process of forming a consideration set and that of making a final decision as two independent decisions, a large variety was more likely to be preferred. However, if those two stages were considered jointly, whereby they only focused on the final outcome, a high level of variety might increase decision difficulty and hence, lead to less satisfaction. It could be the case that in some decision contexts consumers who have to make a decision regularly eventually do not benefit from the variety introduced by different categories within a product domain. This may, consequently, motivate them to put more effort into seeking for variety not only within the product domain, but also beyond it. Menon and Kahn (1995) proposed that presenting product categories from different domains, yet within the product class, might reduce the need for variety and yield positive effect on consumer satisfaction.

Huffman and Kahn (1998) demonstrated that consumer satisfaction could be influenced by how the information about a product is presented. That is, they found that when options information was presented by attributes rather than by product, their subjects were more satisfied with the information gathering process and perceived the choice set as less complex in the case that a large number of options were available. However, when the set of options was small, displaying information by product was preferred. Furthermore, they elicited that the level of complexity of choice may be influenced by consumers' knowledge of the products, whereby those with a high degree of expertise may be more satisfied when the options variety is presented by product rather than by attributes. This implies that attribute-based presentation might increase the decision satisfaction by reducing the complexity of the decision process from a large set of options.

In the online information-seeking context, the search results are typically displayed on a list that is ordered by a single ranking algorithm. This search engine design may raise issues when users attempt to seek for results from various aspects in that the set of available options does not support their needs due to the lack of variety. Under such

circumstances, it would appear that users often adapt their search strategies in such a way that more effort is spent in the search process, which consequently leads to less satisfaction with the search outcome. In addition, if the adapted strategy involves browsing deeper through the ordered results list this may lead to the low quality results (Diehl, 2005). In sum, regarding variety seeking behaviour, it is important to investigate how search engine designs could be improved to support users by presenting search results in such a way that helps them to distinguish the differences between options.

### **2.7 SERENDIPITY**

Pleasure associated with a surprise regarding a consumption experience and an unexpected positive outcome is found to be the experiential bases of high consumer satisfaction (Westbrook & Oliver, 1991; Oliver et al., 1997). In a similar way, it was anticipated that such a surprising experience might increase user satisfaction in search engine use. That is, since search engines have become the basic tool for information-seeking activities, in some situations the occurrence of a surprising or accidental discovery or, so called, “serendipity” could be exceptionally useful to the users, thus increasing their search satisfaction. The examination of literature in this section is aimed at providing the contextual framework for the exploration of chance encountering and serendipitous experiences, as carried out in Chapter 4, as well as for the primary investigation of these in Chapter 5.

Van (1994) defined serendipity as “the art of making an unsought finding”, where the term “finding” refers to various types, e.g. supported hypotheses, observations, ideas, facts, insights or relations, that bring to the fore something new and true (e.g. the discovery of X-ray in science), something new and useful (e.g. the invention of hydraulic constructions technology), or something new and fascinating (e.g. the blue period of Picasso). Furthermore, it was noted that serendipity can easily become underestimated and even denied when the person who comes across it has not got a prepared mind to capture an unexpected incidence. For, as Roberts (1989) explained, an accident becomes discovery because of sagacity of the person who encounters it. Sagacity refers to curiosity and perception, which can be encouraged and developed. For example, students can be encouraged to be flexible in their thinking and interpretation. In addition, to be able to discover, careful and intensive study is the key for preparing the mind.

---

A study by Erdelez (1997) revealed that individuals have different perceptions of their information-encountering experiences, regarding which four groups of respondents were identified. Firstly, *super-encounterers*, refers to those who often find useful and important information, which they rely on occurring as a part of their information seeking process. Secondly, *encounterers* are those who regularly experience and are aware of information encountering, but do not regard it as a part of their information seeking behaviour. The third group, *occasional encounterers*, are those who come across unexpected information randomly, which they perceive as luck. Finally, *nonencounterers*, are those who very infrequently encounter such an unpredicted experience. This author's findings show that individuals have different perceptions about encountering, which may affect the way they react to unexpected circumstances, such that some might try to draw connections between pieces of information that others fail to see, thereby obtaining richer outcomes. This is consistent with an empirical study of interdisciplinary researchers' experiences by Makri and Blandford (2012a). The results from their semi-structured critical incident interviews revealed that different people have a different understanding of serendipity and these understandings tend to change based on encounter experiences. A rough process model of serendipity was proposed (Makri & Blandford, 2012a), which described a chain of serendipitous experience such that a potential connection (a moment of insight) is made between an information need and an unexpected source, which on further exploration indeed proves valuable. Moreover, these authors also defined serendipitous experience as a multi-dimensional conceptual space, which is called "serendipitous space" (Makri & Blandford, 2012b). This space refers to a variety of possible experiences of serendipity that shared three common elements, namely, an amount of unexpectedness, an amount of insight, and a valuable unanticipated outcome.

It has been argued that a serendipity experience can occur when two key components are met: interpersonal, i.e. opportunity or chance, and intrapersonal, i.e. sagacity or a prepared mind (Roberts, 1989; Van, 1994; Rosenman, 1988). Some researchers have investigated the circumstances under which such an experience happens and have concluded that models aimed at enhancing serendipity should do so via the interpersonal component, while others have claimed that providing guidance for improving the intrapersonal element increasing the chances of such an occurrence. Hence, several conceptual frameworks and models have been proposed regarding both individuals'



thinking processes and information-seeking system designs improvement to increase the chances of encountering serendipity (Bates, 1979; Björneborn, 2010; Nutevall & Ryder, 2010; Makri et al., 2014; Campos & Figueiredo, 2001; Hauff & Houben, 2012; Iaquina et al., 2008; Tintarev et al., 2013).

Work by André et al. (2009b) concentrated on exploring the ways to enhance searcher perceptions of information encountering in order to increase the probability of coming across serendipity. They argued that discovery can never happen truly by accident, for knowledge and the correct attitude are needed in order to recognise insights when presented. Consequently, they identified two essential aspects to encountering serendipity: first, the accidental finding of something unexpected and second the ability to assess effectively the discovery found by drawing a connection concerning prior knowledge or perceptions.

The relationship between serendipity and information seeking behaviour is founded on the concept of browsing (Foster & Ford, 2003), that is, the basic features of browsing are the underlying components that lead to a serendipitous experience. Those features were said to be the act of scanning, the presence or absence of purpose, the specificity of search outcomes or goals, and the knowledge about the resources and object sought. Furthermore, it was suggested that a specific goal or purpose might reduce the chance of serendipity encountering, whereby searchers tended not to explore interesting information that was not relevant to their goal (André et al., 2009a; McCay-Peet & Toms, 2010; Toms & McCay-Peet, 2009).

With the aim of understanding how serendipity can be facilitated through the use of technology, McCay-Peet and Toms (2010) interviewed ten historians about their experiences of serendipity. The findings revealed that serendipitous experiences occurred during active learning, specifically in exploratory searching and social networking such that the expertise of colleagues helped them make associations with the information provided. However, recognising serendipity was not always instant, for in some cases the connections emerged only after the current piece of information at hand became related to that found irrelevant in the past. Further, in the context of information seeking via a digital library, it was proposed that the systems should support the users' information acquisition in three cases (Toms, 2000). The first is when the users search for a well-defined and/or known target. Secondly, the systems should support the

---

situation when the target of a search cannot be fully described, but its existence is known. Finally, the system should be able to help users come across an unexpected object or serendipitous discovery.

Serendipity is perceived to be less likely to occur during information seeking via an online environment, because of the quick retrieval of targeted information through keyword searches (Martin & Quan-Haase, 2014). However, André et al. (2009a) argued that in such an environment search algorithms could present searchers with the opportunity for serendipitous experiences, such that some of search results retrieved could be less relevant, but still interesting. In their study, participants' judgements on search results and their past search query logs were investigated, with them being asked to rate each of the search results returned by the search engine based on its relevance and interestingness. The search query logs were used to measure the popularity of queries, the number and diversity of results for each query, as well as user interactions with the search outcomes. The findings revealed that many of the searches people performed returned interesting results that were not directly relevant to the search target. Moreover, personalised configuration appeared to provide interesting results in addition to relevant ones. It is worth noting that in this study serendipity was defined as coming across something interesting that could be related or unrelated to the search goal. However, in real-world situations it may be the case that such results would not be assessed by searchers who rely on a ranking algorithm.

McCay-Peet and Toms (2011) conducted a study investigating serendipitous experience through the use of Wikipedia. The volunteers were asked to explore anything they wished on Wikipedia for twenty minutes so as to familiarise themselves with the system before performing the assigned tasks. After completing each task, they were asked questions concerning their found information and search experience. Five main dimensions of serendipity supported by Wikipedia were elicited from the responses, i.e. it enables connections, introduces the unexpected, presents variety, triggers divergence, and induces curiosity. However, it could not be confirmed that such support would apply across all systems in the digital environment. In terms of search success, it was found that by using a digital library some scientists had missed necessary information by concentrating on a small number of resources in the domain that they were familiar with (Cooksey, 2004). Consequently, the researcher proposed that it is vital for digital libraries to provide support beyond simply accessing digital resources in order to

increase the chance of serendipity encountering. In particular, it was recommended that some guidance should be given in order to help users draw connections between their information needs while searching. Since search engines play a vital role in facilitating information-seeking activities, a better understanding of serendipity encountering in the use of search engines is important so as to support users' needs and their satisfaction in this context.

### **2.8 CHAPTER SUMMARY**

This chapter has provided an overview of the literature relating to concerns with the effects of the number of search results returned and their ordering regarding user behaviour and satisfaction. It has been shown that the number of options and their ordering potentially influence the strategies used in information-seeking processes in an online environment. Firstly, high numbers of search results returned from a search engine is anticipated to lead to choice overload in that users feel less satisfied with their selected options when choosing from an extensive set of returns. Secondly, the ordering of search results may influence users' variety seeking behaviour and the chance of encountering a serendipitous experience during search engine use. The following chapters investigate these issues and discuss the findings in detail. In the next chapter, the investigation begins by exploring the choice overload effect in search engine use.

## CHAPTER 3

# CHOICE OVERLOAD IN SEARCH ENGINE USE

### 3.1 CHAPTER OVERVIEW

This chapter presents a series of three experiments that were conducted to investigate the choice overload effect in search engine use. These experiments were developed to explore the effect in an online environment that closely models real search-engine use, with search behaviour and satisfaction regarding the selected options being investigated. The first concerns the exploration of the choice overload effect in a controlled environment, where the number of options returned by a search engine was manipulated (**RQ1a - What is the effect of choice overload in search engine use?**). The results from this experiment inspired the investigation undertaken in the followed-up experiments, by providing evidence that selecting from a large choice set yields a positive effect, which is inconsistent with previous research.

The literature reviewed in Chapter 2 demonstrated that the behaviour in search engine usage was influenced by ranking bias. Aiming to identify the potential moderators of the choice overload effect (**RQ1b – What are the moderators of choice overload in search engine use?**), the second experiment was conducted by manipulating the ranking algorithm used to order search returns. The results from the second experiment replicate the first, which showed the advantage of a large set of options. The findings also confirm that the perception on ranking of search returns has an influence on the search and selection behaviour as well as affecting the satisfaction with the selected options.

The third experiment investigated further the role of ordering of search returns (**RQ2 – What are the effects of the ordering of search results on the behaviour and satisfaction of search engine users?**), by examining the effect of time limitation on choice behaviour and satisfaction judgement. Evidence emerged that time pressure is an important determinant of choice overload, which provides support for process-product leakage theory. The results reveal that the satisfaction with selected options is affected by the number of options interacted with the time pressure. Moreover, the ordering of search results affects selection processes and satisfaction judgements.

The main analyses of this series of experiments were based on quantitative data gathered via 10-point Likert scales. Additionally, information on self-reported search strategies was collected through a questionnaire and was gathered and analysed in order to gain insights into the effect of choice overload and its moderators.

---

## 3.2 INTRODUCTION

Users of current web search engines are typically presented with a large number of returns after each query, ordered according to some obscure algorithm that makes it likely, to some extent, that valuable hits will appear early in the list. (For example, PageRank (Page et al., 1999), a proprietary ranking algorithm used by Google search in determining the relative importance of a web page. The exact way this algorithm works is unknown outside Google, but it can be assumed that the level of importance incorporates an assumption that the more links a web page receives, the more important the web page is.) The length of the list means that users must choose returns to open and inspect without consideration of most of the alternatives. As noted by Oulasvirta, Hukkinen and Schwartz (2009), in the study reviewed in Chapter 2, this situation may lead to “choice overload”, i.e. negative psychological effects of being asked to choose from a large set of options (e.g., Iyengar & Lepper, 2000; Arunachalam et al., 2009; Fasolo et al., 2007; Iyengar et al., 2006; Loewenstein, 1999, Schwartz, 2004).

However, as we have seen, the empirical evidence concerning the choice overload effect is somewhat contradictory, with as many studies reporting positive effects of large choice sets as those that have been consistent with the findings of negative consequences of Iyengar and Lepper (2000). This inconsistency could be attributed to the set of options controlled in the studies not being “large enough” to create an overload effect in some contexts. Moreover, this situation is blurred further by there being no agreement on what set is “large enough” such that the choice overload effect will occur.

Returning to the single published study that investigated a choice overload effect in a search engine use context (Oulasvirta et al., 2009), it appears that several aspects of the experimental design of this study limit the generalisation of the conclusions to real search engine use. That is, the study of Oulasvirta et al. used paper-based materials to model aspects of a typical search engine scenario, which could thus be considered as only a proxy investigation for real-world circumstances and hence, possibly lacking authentic outcomes. Moreover, the strong time constraints employed, i.e. thirty seconds per search task, would have been likely to affect satisfaction with the search outcomes.

Bearing these issues in mind, the aim for this research was to develop Oulasvirta et al.’s work with an online experiment that would more closely model the use of real search

engines, and which would allow participants to choose from the sets of snippets and consult web pages they selected before rating their satisfaction with those web pages.

### **3.3 EXPERIMENT 1: NUMBER OF OPTIONS AND ITERATION**

#### **3.3.1 METHOD**

##### **3.3.1.1 DESIGN**

This experiment aimed to replicate the main features of Oulasvirta et al. (2009), but online, with participants doing real-time Google searching, specifying their own search terms and choosing a webpage to view before making any judgements. Only the Problem task type was used. Two separate sets of 10 questions were developed and utilised, but after preliminary analysis the question set was not treated as an independent variable in the main analysis. (Example problems are given in the different method sections below and the full set is available in Appendix B.2)

This experiment involved manipulating the number of options returned by Google: six in a smaller set size condition (on a single Search Engine Results Page (SERP), see Figure 3.1) and twenty-four in a large set size (on four SERPs, each reached by pressing the page number button at the bottom of each page, see Figure 3.2). Individual search terms were created by the participants themselves, with half being allowed to iterate or change their search terms after inspecting the returns, whereas the other half had to select a webpage after the first search. Both independent variables were manipulated between subjects, giving a 2x2, Set Size (6 vs. 24) x Iteration (iteration vs. no-iteration) between-subjects design. For each of the 10 questions in their question set, the participants were required to search for information by using Google then choose a single webpage to explore from the SERPs. If they opened a selected web page, they were not allowed to reject it and search for another, which meant their decision had to be made entirely from the snippets. After each webpage had been chosen, they inspected the page and then rated their satisfaction with it, their trust in its reliability and their judgement of its relevance for the particular assigned task.

##### **3.3.1.2 PARTICIPANTS**

The thirty-two participants were students and academic staff from the University of Bath, 16 males and 16 females. Their ages ranged between 23 and 45, with the average

being 29.9 years. All reported regularly using Google in their everyday lives. To incentivise their performance, they were told that the quality of the selected websites would be evaluated in order to allocate a cash prize and after finishing the experiment, each was given a box of chocolates as a token of appreciation for their participation.

### 3.3.1.3 MATERIALS

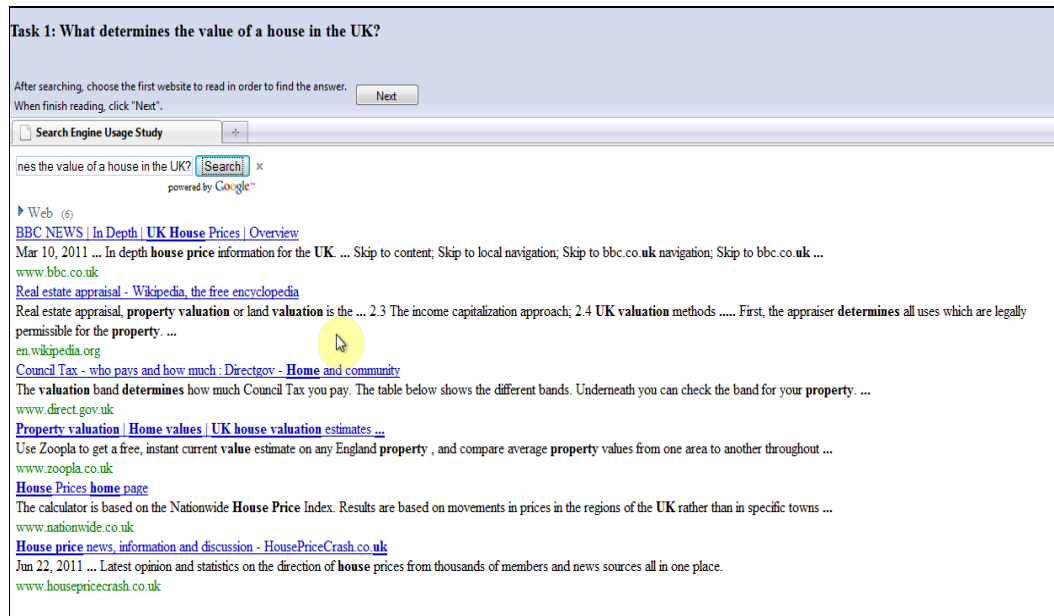


Figure 3.1 Experiment 1: An example of a computer screen presenting six options

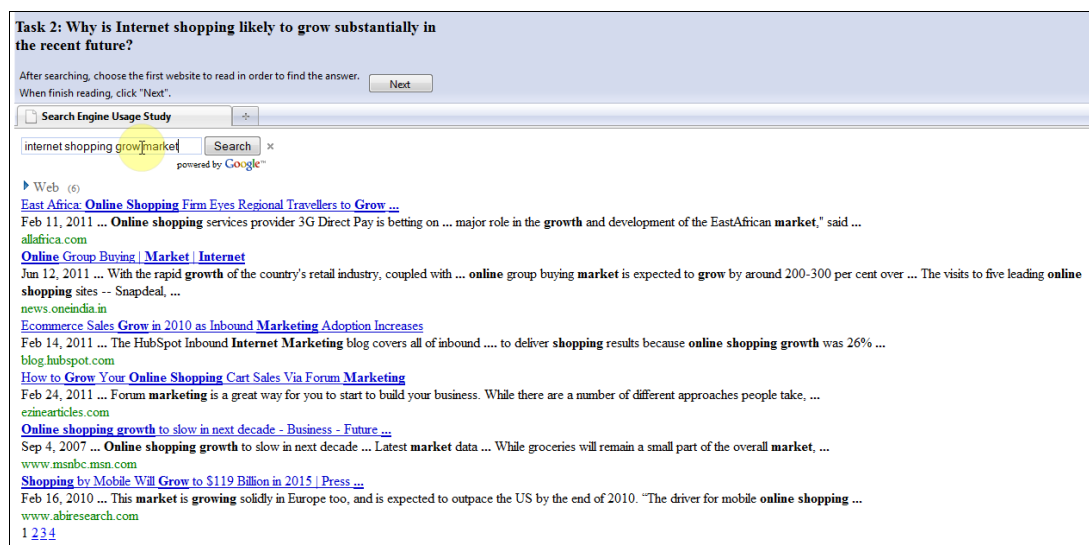


Figure 3.2 Experiment 1: An example of a computer screen presenting twenty-four options (on four SERPs)



HCI Browser (Capra, 2010), an open-source extension to Mozilla Firefox 3, was modified to collect data and to guide the participants through their tasks. Figure 3.1 and Figure 3.2 show examples of a search interface presenting six and 24 options, respectively. In addition, the interaction on the computer screen was recorded. Google was used by all participants to complete each task: SERPs were altered using Google's API.

After each task was completed the participants rated their selected website on 10-point Likert scales. Additionally, there was a single open-ended question inviting a brief typed answer to a request for them to justify their satisfaction rating. The question protocol was as follows:

- How satisfied are you with your selected page?
- Why? [open-ended question]
- To what extent was the information provided by the website relevant to your task
- To what extent do you trust the information provided by the website.

After the set of 10 search tasks was completed, a final questionnaire asked the participants about the strategy they used to choose web pages (an open-ended question), overall satisfaction for those they selected and overall satisfaction for the list of options that the search engine returned (10-point Likert scales).

### **3.3.1.4 PROCEDURE**

The participants completed the experiment individually in a laboratory. The experiment was divided into two sub-sessions: training and experimental. On arrival at the laboratory, they were introduced to the HCI Browser interface that would guide them through the experiment. They were instructed that their general task would be to search for web pages that provided valuable information for a series of separate questions or tasks. They then completed one sample task during the training session, with the experimenter available to offer guidance as required. The question for this training task was "Why did ancient Egyptians mummify their dead?" and the participants undertook this task with the same interface and iteration constraint as would apply for their experimental session. After each participant had confirmed that they understood the general task and the user interface, the experimental session started. Each task question was presented on the screen at the top of the browser window and remained there during the search and the subsequent ratings, with no time limit being specified. In a debriefing

---

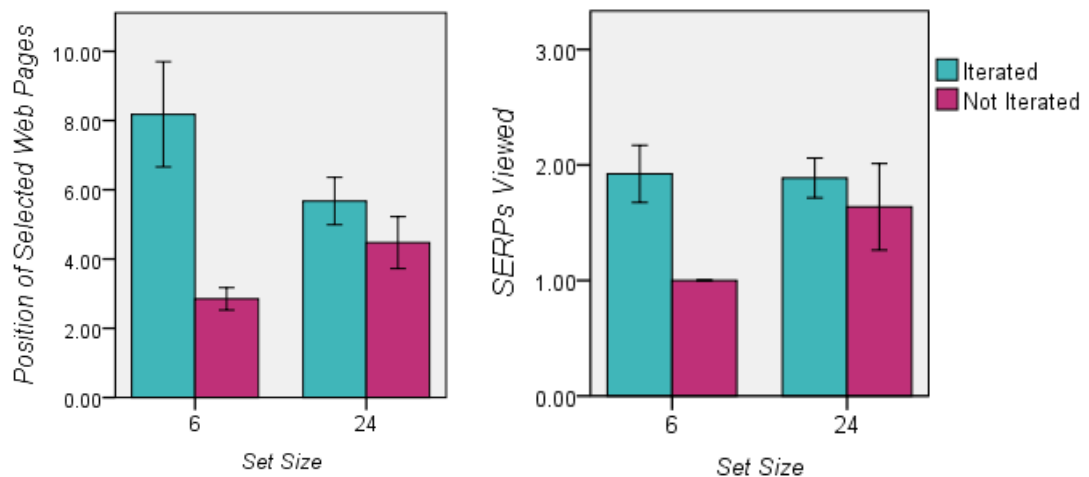
session, the purpose of the experiment was explained to the participants, and they were allowed to ask any question related to the study.

### 3.3.2 RESULTS

The performance of the participants answering question groups A versus B was compared using t-tests and no significant difference was found between these groups for any of the dependent variables considered in the main analysis below. Therefore, the distinction between question groups was not further considered. The analyses reported below are 2 x 2, Set Size x Iteration between-subjects ANOVAs.

#### 3.3.2.1 POSITION OF THE SELECTED WEB PAGES

Figure 3.3 (left) shows the mean ordinal position of the selected web pages, where it can be seen that most were chosen from the top of the result lists in both the small set and the large set of options. Indeed there was no significant effect of Set Size on the ordinal position of the selected web pages ( $F(1,28) = 0.2, p = 0.64$ ). However, those participants who could perform the iteration quite often changed their search term (see right hand panel, the mean of c. 2 SERPs viewed in the small set size means a mean of one iteration, i.e. two search-terms used). A simple main effect within the iteration condition showed a significant difference in selected webpage positions between the small and the large set sizes ( $F(1,28) = 12.3, p = 0.002, \eta^2_p = 0.31$ ).



**Figure 3.3 Experiment 1: Mean position of the selected web pages (left), and number of SERPs viewed before selecting (right). Error bars show standard errors.**

Note that in the iterated condition we accumulated the selected position and the SERPs viewed across iterations, e.g. choosing the second snippet after one iteration in small set size would yield a position of 8 and SERPs viewed of 2.

Figure 3.3 (right) shows the mean number of SERPs that the participants viewed (out of a maximum of 4 separate SERPs for 24 items) before selecting a webpage. Those who were under the large set size condition only sometimes viewed beyond the first screen, and indeed there was no significant effect on the number of SERPs viewed across set size conditions ( $F(1,28) = 1.5, p = 0.22$ ). The number of SERPs that they viewed was also updated by iteration and hence, as expected, the number of SERPs viewed between the iteration and non-iteration conditions were significantly different ( $F(1,28) = 5.9, p = 0.021, \eta^2_p = 0.18$ ). Moreover, participants in the small set condition were more likely to iterate ( $M = 1.92, SD = 0.69$ ) than those with the large set ( $M = 1.26, SD = 0.36$ ). Independent t-test (among the 16 participants who could iterate) established this effect as significant:  $t(14) = 2.37, p = 0.038$ .

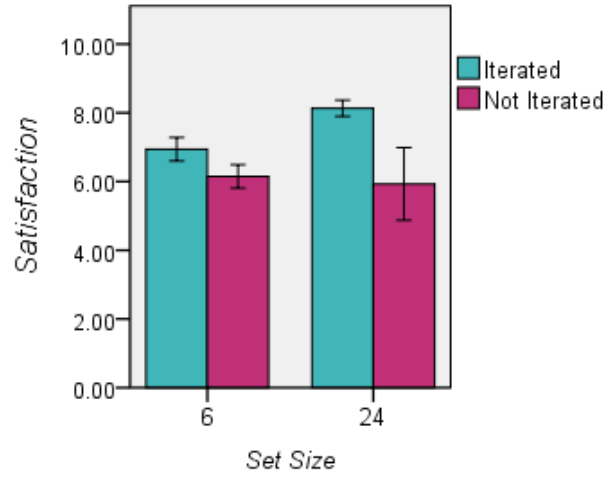
### 3.3.2.2 PERFORMANCE TIME

Neither of the independent variables had a significant effect on the time to select or view a web page, nor was there an interaction effect. The overall mean time to select a web page was 126 seconds; the overall mean time to view a web page before rating it was 34 seconds.

### 3.3.2.3 RATING OF THE SELECTED WEB PAGES

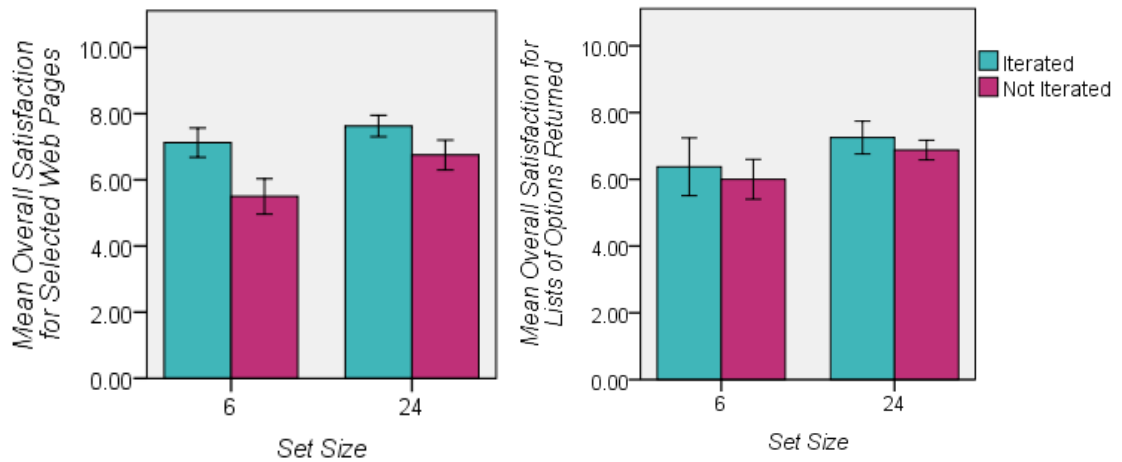
Figure 3.4 displays the participants' mean satisfaction with their chosen websites according to experimental condition. ANOVA confirmed that those participants in the large Set Size were marginally reliably more satisfied with their selected web pages ( $F(1,28) = 3.9, p = 0.056, \eta^2_p = 0.12$ ). Moreover, those who could iterate their search terms were reliably more satisfied with their selected web pages ( $F(1,28) = 9.9, p = 0.004, \eta^2_p = 0.26$ ). However, there was no significant interaction between Set Size and Iteration ( $F(1,28) < 1$ ). No significant effects were found for Trust judgements (Trust was significantly correlated with Satisfaction,  $r = 0.68, p < 0.001$ ). In addition, Relevance judgements behaved very similarly to Satisfaction ones, and these judgements were very strongly correlated with each other ( $r = 0.83, p < 0.001$ ). The only significant effect was a main effect of Iteration on judgements of relevance, ( $F(1,28) = 4.3, p = .046, \eta^2_p =$

0.14). All other main effects and interactions failed to reach significance, with all  $F$ 's being close to 1.



**Figure 3.4 Experiment 1: Mean satisfaction with the selected web pages.**  
Error bars show standard errors.

### 3.3.2.4 OVERALL RATING OF THE SELECTED WEB PAGES



**Figure 3.5 Experiment 1: Mean overall satisfaction for the selected web pages (left), and mean overall satisfaction for the lists of options returned (right). Error bars show standard errors.**

Figure 3.5 shows participants' satisfaction ratings at the end after all the tasks were completed. They were asked to rate their overall satisfaction for the selected web pages and for the lists of options returned at the end of the experiment, which were presumably reflecting memory for the overall experience across 10 tasks. They were marginally more satisfied with the selected web pages in the large option set condition than were participants under the small set condition ( $F(1,28) = 3.8$ ,  $p = 0.059$ ,  $\eta_p^2 = 0.12$ ). In

addition, those who could iterate their search term were reliably more satisfied with their selected web pages ( $F(1,28) = 7.9, p = 0.009, \eta^2_p = 0.22$ ). However, the interaction effect between Set Size and Iteration conditions was non-significant ( $F(1,28) < 1$ ) and there was also no effect of Set Size on Satisfaction with the lists of options returned ( $F(1,28) = 2.1, p = 0.155$ ).

### 3.3.2.5 ITERATED SEARCH TERMS

As previously mentioned, participants who can perform iterations quite often change their search terms and in this subsection the changes of these during iterative searching are described according to three categories. In the first, some participants made iterations by substituting a word in their search term(s) with its synonym or another related word. For example:

[Participant #7, Small Set]

Task question: Why is China famous for pottery?

Search term: *reason* china famous pottery

1<sup>st</sup> iteration: *characteristic* china famous pottery

[Participant #12, Large Set]

Task question: What determines the value of a house in the UK?

Search term: UK house *price*

1<sup>st</sup> iteration: UK house *value*

Secondly, the changing of search terms was performed in order to increase or decrease the specificity of the search results returned. For example:

[Participant #4, Small Set]

Task question: Why did The King's Speech win the best director Oscar for Tom Hooper?

Search term: king's speech Oscar

1<sup>st</sup> iteration: king's speech Oscar best director

2<sup>nd</sup> iteration: why did king's speech win Oscar best director

[Participant #8, Small Set]

Task question: Why is Internet shopping likely to grow substantially in the recent future?

---

Search term: why is Internet shopping likely to grow substantially in the recent future? (Note: the task question was used as the search term)

1<sup>st</sup> iteration: Internet shopping

2<sup>nd</sup> iteration: online shopping

3<sup>rd</sup> iteration: online shopping reason

Finally, the iterations were performed by using search tactics, i.e. using the asterisk sign (\*) that is a wildcard character referred to “LIKE” expression, which means matching zero or more characters<sup>3</sup>. The following example shows that the participant used all iteration categories identified in order to perform the search for one task.

[Participant #3, Small Set]

Task question: What determines the value of a house in the UK?

Search term: value of house uk

1<sup>st</sup> iteration: value of house uk how determined

2<sup>nd</sup> iteration: how to work out value of house uk

3<sup>rd</sup> iteration: determines value house uk

4<sup>th</sup> iteration: determines value house England

5<sup>th</sup> iteration: what determin\* value house England

6<sup>th</sup> iteration: factors determin\* value house England

The above shows that the participants used different strategies for changing search terms in order to retrieve different sets of results.

### **3.3.2.6 STRATEGIES FOR CHOOSING A WEB PAGE**

The qualitative data about participants’ strategies for choosing a web page were analysed by using thematic analysis (Braun & Clarke, 2006). Their responses were relatively short in length, ranging from one sentence to a short paragraph. The analysis was aimed at exploring the search strategies in the given experimental conditions and inductively eliciting factors relevant to those strategies. The transcripts were read twice and the initial themes were identified by the researcher. The transcripts were then given to an independent coder in order to be checked for the presence or absence of the identified themes. However, the results reported were based solely on the researcher’s coding and hence, inter-coder reliability was not performed. Four relevance themes were identified:

---

<sup>3</sup> See example: <http://www.computerhope.com/jargon/w/wildcard.htm>

*keyword matching; reliability of information sources; ranking of option; and options from alternative queries.*

First, *keyword matching* pertained to the strategy of choosing a web page based on the title and snippet content. This was used by all the participants as the base-strategy to verify the relevance of options. Some tended to match the keyword from the task question with the titles of options in the result lists, while others read through snippets:

[Participant #1, Small set] *“Reading description of each link and make judgement based on perceived relevance of each link to the question”*

[Participant #7, Small set] *“Number of keyword reappearances, page title mainly”*

[Participant #24, Large set] *“The title covers the main words from the questions”*

[Participant #26, Large set] *“Looking at the short text that was outlined below each search result and checking to what extent it indicated that it might answer the question”*

Second, *reliability of information sources* referred to statements mentioning the consideration of well-known or famous web sites, the organisation that provided the web page, and the trustworthy or academic web sites. This showed that the participants used their preference and recognition to judge the reliability of the web site. It seems that although they may not have been familiar with the content related to the task question, they would expect that from a recognised or preferred web site to be reliable:

[Participant #16, Small set] *“Notice whether the page is provided by a famous website like BBC or Wiki”*

[Participant #22, Large set] *“Web address – is it a known page (Wikipedia), is it based on opinion (yahoo/social network), is it biased (certain news sites), is it a balanced reputable source?”*

[Participant #28, Large set] *“I prefer Wiki which can be deemed as a portal for all relevant information. It is a search engine itself in some way. Besides, I prefer famous web sites rather than a website that I never heard of, Surely, the websites, whose URL end with .gov, .org, etc., would be paid more attention.”*

---

Third, *ranking of options* pertained to statements where the participants mentioned only exploring options presented on the first result page or judged relevance of options based on the ranking:

[Participant #16, Small set] *“Since the browser just returns 6 pages, the ranking is not important however if more results would be returned, only the first 10 results will be explored I am afraid.”*

[Participant #18, Large set] *“I always used the first page as there were good enough answers there.”*

Finally, regarding *options from alternative queries*, one participant from the small set condition mentioned iterating the queries to explore more results. This could be influenced by the limit number of options presented in the small result list:

[Participant #6, Small set] *“I would try a few alternate queries as well to see if I could find anything better. Sometimes I would go back to the first set of results.”*

These themes offer insights into the strategies in the search and selection behaviour that provide support for the quantitative analysis. This analysis also uncovered some of the possible factors that may have an impact on search satisfaction.

### **3.3.3 DISCUSSION**

This experiment provided no evidence for a choice overload effect. Indeed, the effect of set size on choice worked in the opposite direction, with a marginally significant greater satisfaction being reported by those participants who chose web pages from a set of 24 returned snippets. This result contrasts with the findings of Oulasvirta et al. (2009), despite the overlap in task context and the identical set sizes in the two experiments. It is contended that some of the issues reviewed above may explain the contrast. First, Oulasvirta et al. asked participants to report satisfaction with a chosen snippet without actually consulting the web page to which it related. This, it seems to us, makes it even more likely that satisfaction judgements will be influenced by the search process (because snippets contain far less information and therefore are less differentiated than are full web pages). Second, these authors imposed a very strict time limit, which again, as argued above, is likely to have resulted in an unsatisfactory decision process in the case of large sets: participants in the large set condition in their experiment were allowed



just over one second per snippet to make their choice. Third, large sets for the current research were paginated, whereas Oulasvirta et al. presented them as a single list and as explained above, this makes the size of the set less salient and more manageable.

The other very important difference in experimental methods used was that the participants for this work could construct their own search terms, and half were allowed to iterate over search terms if dissatisfied with the initial set of returns. It is unsurprising that there was a main effect of iteration on satisfaction judgements, for this shows that the participants were able to judge the quality of web pages from the snippets and hence, were able to improve their search terms if the initial results were disappointing. Moreover, the participants who received a small set of options iterated more than did those with a large set, which suggests that the former condition was more likely to be judged as unsatisfactory, thereby adding additional support to the conclusion that, in this experiment, larger sets of returned snippets were more satisfactory than smaller ones. It is striking that most of the participants under the large option set condition usually selected a webpage from the first page, often without bothering to iterate their search terms or browse options on later result pages. This is consistent with the finding from Pan et al. (2007), which showed that user's choice of the particular web page was mostly based on its position on the results list. It is also in line with our suggestion above that the size of the returned set is not psychologically salient (although of course this cannot explain the advantage we observed for the large set).

### **3.4 EXPERIMENT 2: NUMBER OF OPTIONS AND RANKING**

The results of the first experiment suggest that choice overload may be mitigated or reversed in the case of search engines, and we have argued that an important reason may be the perception that the order of the choice set is a good, if approximate, guide to quality. The second experiment was designed to investigate this effect further.

#### **3.4.1 METHOD**

##### **3.4.1.1 DESIGN**

The participants' expectation about the ranking algorithm was manipulated as a within-subject independent variable. In particular, they were informed that one of three processes was used to rank the returns of a search (although in fact, in all cases the

---

Google ranking was preserved, as explained to them during the post-experiment debriefing):

- *Expert ranking*: Participants were informed that the web pages linked to in the returns-list had been ranked by an expert in the appropriate topic area, according to how well they had answered the question.
- *Novice ranking*: Participants were informed that the web pages had been ranked according to how well they had answered the question, but by someone with no particular knowledge of the field.
- *Random ranking*: The order of the list of links has been randomised.

To make this manipulation plausible, the design of the study was constrained such that pre-specified search terms were used and indicated to participants (in this respect, this second study moves closer to the study by Oulasvirta et al. (2009)) and these terms were abbreviated versions of the problem specification. For example, when the task question was “What determines the cost of living in the UK?” the search term was “cost of living in UK”. Consequently, there was no iteration of search terms in this experiment.

Three separate question sets were used (four questions per set). All the participants received all the three question sets, with each question set being associated with a particular ranking type. Ranking types were assigned to the question sets, and ordered so that each ranking type was associated with each set in an ordinal position equally across the participants. This resulted in, a 2(set-size: 6 vs. 24, between-subjects) x 3(ranking type: expert vs. novice vs. random, within-subject) mixed design.

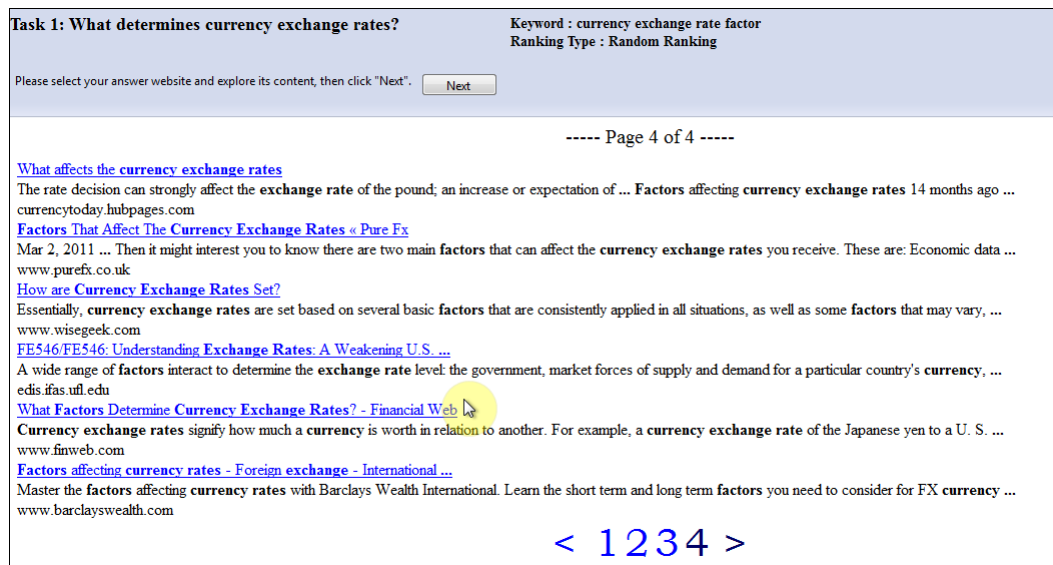
For each of 12 tasks, the participants were required to select and open a single webpage that provided valuable information from the list of search results. After each webpage was selected, they were required to assess it and rate their satisfaction level in terms of their trust in its content, the relevance of the page for the task, and their familiarity with the task question. This last question was added as a check that one of the main necessary factors for choice overload was not compromised by participants’ prior experience. One final change to the procedure was a minor redesign to the user interface, so that the page numbers of search returns beyond the first page in the large set was more salient (the page number was increased in size and each returns page was additionally labelled with “----- Page <Nth> of 4 -----” (see Figure 3.6).

### 3.4.1.2 PARTICIPANTS

Twenty-four participants from the University of Bath were recruited via an online advertisement and posters on notice boards, 12 males and 12 females. No one had taken part in Experiment 1. They were aged between 19 and 39, with an average being 27.9 years. Each was paid five pounds in cash for their participation. They were instructed that the quality of the selected webpages would be evaluated in order to allocate the prizes. That is, three cash prizes were used to motivate them to search for high quality webpages.

### 3.4.1.3 MATERIALS

Similar to the first experiment, the HCI Browser was modified to collect data and to guide the participants through their task. That is it was modified, as above, to make the size of larger sets more salient. Figure 3.6 shows an example of the search interface presenting 24 options in the Random ranking condition. After a webpage was selected for each task the participants were asked to rate the webpage on 10-point Likert Scales, indicating satisfaction, then relevance and finally trust, as in Experiment 1. In addition a question about the familiarity was added, i.e. “To what extent is the task question familiar to you?”



**Figure 3.6 Experiment 2: An example of a computer screen presenting twenty-four options (on four SERPs) in the Random ranking condition**

---

After the participants had finished one set of tasks for each ranking type, a questionnaire asked them about the strategy they had used to choose the web pages (an open-ended question), their overall satisfaction for their selected pages and their overall satisfaction regarding the lists of pages that the search engine returned (10-point Likert scales). In addition, in order to obtain rich qualitative data about the strategy used, they were encouraged to describe and justify their strategies in detail.

### **3.3.1.4 PROCEDURE**

The procedure was similar to that of Experiment 1, with the participants completing the experiment individually in a laboratory. The study was also divided into training and experimental sessions. During the former, they were introduced to the HCI Browser interface that would guide them through the experiment and were instructed that their general task would be to search for web pages that provided valuable information for a series of separate questions or tasks. The three types of ranking algorithm were described to them in written instructions and any questions were answered by the experimenter. During the main part of the experiment, participants were informed by a screen message about the type of ranking used for each task set before seeing the task questions. In addition, the browser window displayed the task question, the search term and the ranking type at the top of the window. This information remained on screen throughout the search task and the subsequent ratings. No time limit was specified. In a debriefing session, the purpose of the experiment and the deception concerning the descriptions of ordering of search returns were explained to the participants, and they were allowed to ask any question related to the study.

### **3.4.2 RESULTS**

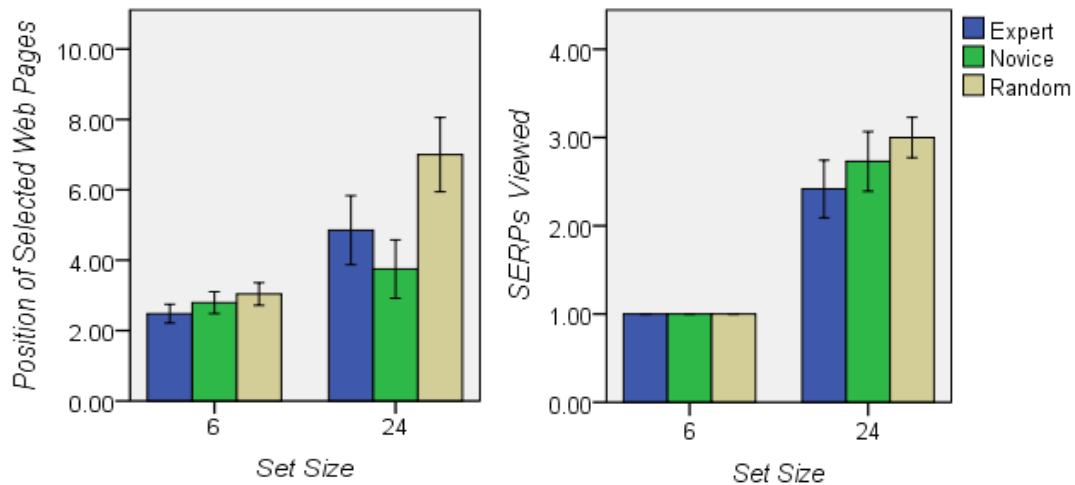
The main analyses reported below are 2x3, Set Size (6 vs. 24) x Ranking Type (expert vs. novice vs. random) mixed ANOVAs.

#### **3.4.2.1 POSITION OF THE SELECTED WEB PAGES**

Figure 3.7 (left) shows the mean ordinal position of the selected web pages. The results indicate that the participants selected web pages from further down the list when they were provided with the large set of options. The main effect of Set Size on the ordinal position of the selected web pages was significant ( $F(1,22) = 11.2, p = 0.003, \eta^2_p = 0.34$ ). In the Random ranking condition, the selected web pages were located considerably

further down the search result list. The main effect of Ranking Type was significant ( $F(2,21) = 4.9, p = 0.018, \eta^2_p = 0.32$ ). Further, the interaction effect between Ranking Type and Set Size was found to be significantly different ( $F(2,21) = 4.1, p = 0.031, \eta^2_p = 0.28$ ).

Although it remained the case that in the large set condition the majority of the selected web pages were located in the first SERP (Figure 3.7, left), most of the participants checked the search results provided in other SERPs before selecting their preferred web page (Figure 3.7, right). According to a one-sample t-test, the number of SERPs viewed in the large set size was significantly greater than 1 ( $t(11) = 6.4, p < 0.001$ ). The main effect of Ranking Type on the number of SERPs viewed in the large-set condition was non-significant ( $F(2,22) = 2.98, p = 0.07$ ).



**Figure 3.7 Experiment 2: Mean position of the selected web pages (left), and number of SERPs viewed before selecting (right). Error bars show standard errors.**

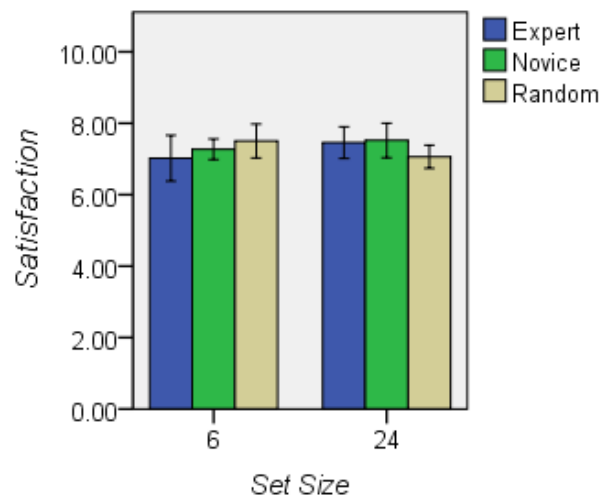
### 3.4.2.2 PERFORMANCE TIME

Neither of the independent variables had a significant effect on the time to select or view a webpage, nor was there an interaction effect. The overall mean time to select a webpage was 67 seconds; the overall mean time to view a webpage before rating it was 35 seconds.

### 3.4.2.3 RATING OF THE SELECTED WEB PAGES

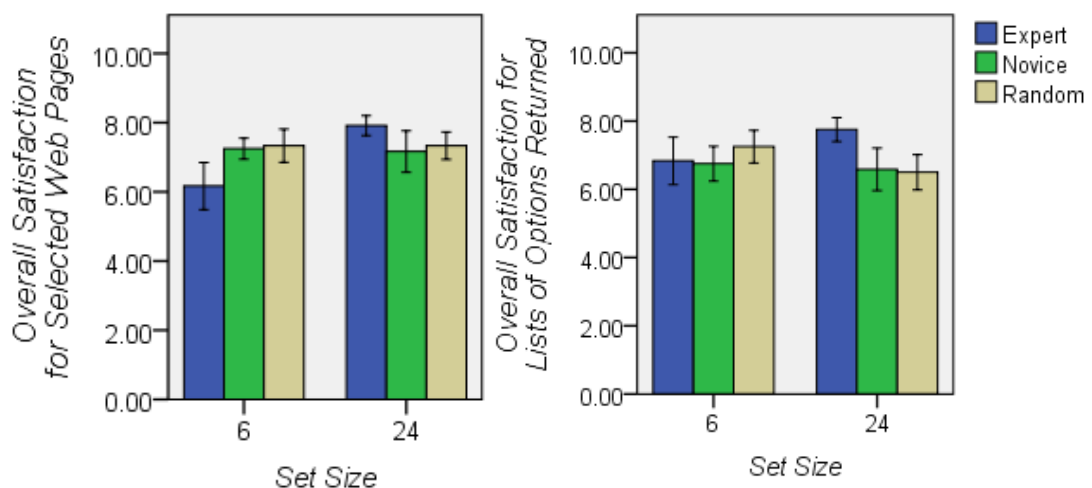
Figure 3.8 shows the participants' rated satisfaction with chosen web pages. There was no significant main effect of ranking type or set size and no interaction effect (all  $F$ s < 1). Similarly, there were no main effects or interactions on judgements of Trust,

Relevance, and Familiarity, with all  $F$ s being close to 1. Trust was significantly correlated with Satisfaction ( $r = 0.52, p < 0.001$ ). Relevance judgement was even more strongly correlated with Satisfaction ( $r = 0.87, p < 0.001$ ). Familiarity judgements averaged around 5, suggesting that the participants were not over-familiar with the choices they were asked to make. Finally, the correlation between Familiarity and Satisfaction was non-significant ( $r = 0.19$ ).



**Figure 3.8 Experiment 2: Mean satisfaction with the selected web pages.**  
Error bars show standard errors.

#### 3.4.2.4 OVERALL RATING OF THE SELECTED WEB PAGES



**Figure 3.9 Experiment 2: Mean overall satisfaction for the selected web pages (left), and mean overall satisfaction for the lists of options returned (right). Error bars show standard errors.**

Figure 3.9 shows the participants' satisfaction ratings at the end of each task-block (i.e. after each ranking type). There was no significant main effect of Ranking Type or of Set Size on overall satisfaction for the selected web pages, but the interaction effect was reliable ( $F(2,21) = 3.9, p = 0.036, \eta^2_p = 0.27$ ). The interaction effect between Ranking Type and Set Size on overall satisfaction for the lists of options returned by the search engine was non-significant ( $F(2,21) = 1.98, p = 0.16$ ). Neither Ranking Type nor Set Size yielded the main effect on overall satisfaction for the lists of options returned ( $F$ s close to 1), which showed that the participants did not perceive any difference between the result lists from different ranking algorithms. This was reliably consistent with the fact that SEPRs in each task across all ranking conditions were identical.

#### 3.4.2.5 STRATEGIES FOR CHOOSING A WEB PAGE

The analysis of the qualitative data involved the same approach as Experiment 1. Initial codes were developed and resulted in three main themes: *ranking reliance*; *keyword matching*; and *reliability of information sources*. The findings provided insights into the participants' perception and level of reliance on the ranking algorithm. It emerged that most of the participants employed multiple strategies in sequence to eliminate the relevant options. First, *ranking reliance* pertained the statements where the participants mentioned that they selected the option based on its position ranked by the reliable ranking algorithm. This strategy was the primary one used by most of those in the large set condition. In particular, most mentioned that they mainly paid attention to options on the first results page, especially when they selected options in the Expert ranking condition:

[Participant #1, Large set, Expert ranking] “*As it is expert ranked, I chose from the first 3 items.*”

[Participant #3, Large set, Novice ranking] “*As it is novice ranked, I didn't read through the small print as carefully as I did for the randomly ranked list but did read throughout to make sure that the information provided by the providers was relevant to what I was looking for.*”

[Participant #12, Small set, Random ranking] “*Scanning all presented options, ignore the ranking.*”

---

[Participant #21, Large set, Expert ranking] *“Generally didn’t check pages other than Page 1 because I thought the expert ranking would mean poorer results were on other pages.”*

Second, *keyword matching* referred to statements reporting that the titles and snippets of search results contained keywords from the task question. This was the strategy that was mainly used by the participants in the Novice and Random ranking conditions. In addition, the majority of participants under the small set condition used this strategy when they chose from the result lists that they believed to be ordered by Expert ranking. For example:

[Participant #8, Small set, Expert ranking] *“To see if the content contains the keywords.”*

[Participant #11, Large set, Random ranking] *“Looking at the keywords highlighted in the snippets and determining if it is relevant to the questions.”*

[Participant #21, Large set, Novice ranking] *“Reading the title for appropriateness and then scanning the description to see if the context was relevant.”*

[Participant #24, Small set, Expert ranking] *“Looking for the most relevant words.”*

Finally, *reliability of information sources* pertained to statements commenting on the consideration of credible and reliable web sites or organisations that provided information. Some participants mentioned their concerns about up-to-date information. For example:

[Participant #5, Small set, Novice ranking] *“Trusted website go first then check brief information about my question.”*

[Participant #11, Large set, Novice ranking] *“I tried to look for a web page which is from a well-known organisation or company. And also for certain questions I tried to look for the latest articles so that I would get the latest updates.”*



[Participant #13, Large set, Random ranking] *“Go through all search results and choose the source which seemed most reliable.”*

[Participant #14, Small set, Expert ranking] *“First, I chose those having words most pertinent to the question, then I chose the one that was the most trustable, such as Wikipedia or gov.uk.”*

It is contended that these themes offer insights into search and selection behaviour, in that it has been elicited that strategies employed were based on the perceived efficiency of the ranking algorithm. Generally, the sequence of strategies applied was varying based on the search task questions.

### 3.4.3 DISCUSSION

The participants' selection behaviour confirms that they were influenced by what they believed to be true about the ranking of returns. That is, they did not typically simply accept the Expert's ranking (under that condition, the average ranking for the chosen website is 2 in the small set, and about 4 in the large set). Nevertheless the participants chose from further down the set of returns when they believed the ranking was random, and when they were presented with more returns (i.e. the Set Size is large).

As in Experiment 1, there is no hint of a choice overload effect in the satisfaction ratings of the chosen web pages. This is true even when the participants believed the ranking of returns was random, thus showing that their perception that the rank ordering is helpful is not necessary for explaining the relative positive reaction to large set sizes. (That the rank ordering *is* typically helpful may still be important in this respect.) Participants in the Random, large Set-Size condition certainly processed the size of the set (often inspecting all four SERPS and choosing on average the web page ranked seventh), but this did not lead to significantly less satisfaction. As it happened, these participants were still choosing early items more than late items from the results list, and this in itself may mitigate any negative effects of the large set.

In sum, the outcomes of this experiment have not replicated the satisfaction advantage of a large set size reported in Experiment 1, except when the overall ratings of chosen web pages in a task block are considered under the Expert ranking condition (see Figure 3.9).

---

### 3.5 EXPERIMENT 3: NUMBER OF OPTIONS, RANKING AND TIME PRESSURE

One important difference between our first two experiments and the typical choice overload experiment, especially that of Oulasvirta et al. (2009), is that our participants were not given any time constraints. We have suggested that such constraints may be operational in choice overload contexts in that they lead to dissatisfaction with the choice process, which “leaks” into the judgements regarding the chosen items. The third experiment explores this issue as well as seeking further evidence concerning the importance of the ranking of returns on participants’ search behaviour and their subjective judgements.

#### 3.5.1 METHOD

##### 3.5.1.1 DESIGN

Time limitation was manipulated in this experiment, with the participants being given either 45 seconds or 90 seconds to finish each task. To study further the effect of ranking, participants were provided with the search results that were generated from one of the two ranking algorithms: Google ranking and Random ranking.

- *Google ranking*: All the web pages that are linked to in the returns-list have been ranked by Google.
- *Random ranking*: The order of the list of links has been randomised.

In the Google ranking condition, SERPs were generated by Google and broken links were eliminated from the returns lists in order to reduce noise in the experiment. In the Random ranking condition, the search results in SERPs were those from a Google-ordered list. However, the order of them was in fact random (a single random order was used for each task/set-size).

As in Experiments 1 and 2, the between-subjects independent variable was the set size of a results list returned by the search engine, which was either 6 (on a single SERP) or 24 items (on four separate SERPs, each reached by pressing a page number button at the bottom of each). There were 16 tasks in total, which were divided into four blocks (four tasks per block). Each block had a different combination of ranking order and time limitation. Half the participants performed the two Google blocks before the two Random blocks, whereas for the other half this was reversed. Similarly, half the

participants performed the 90s block before the 45s block for each Ranking, whilst for the rest this was reversed.

After each webpage was selected, the participants assessed it and then rated their satisfaction regarding the selected webpage, their trust in its reliability, its relevance for the task, and their confidence that their selected webpage was the best in the search results list (10-point Likert Scales). The use of a confidence rating was novel in this experiment. Regarding this, a previous study by Heitmann and colleagues examined consumer behaviour in the electronics market context (Heitmann et al., 2007), with their results showing a positive correlation between confidence and satisfaction. It seems to us that it might confirm that the satisfaction regarding the selected options might be affected by the decision making process, e.g. confidence would be low if only poor consideration of alternatives was possible. The additional question was “How confident are you that your selected website is the best in the set you chose from?”

After each session was completed, a questionnaire asked the participants about the strategy they used to choose web pages (an open-ended question), overall satisfaction regarding their selected pages and overall satisfaction in relation to the list of pages that the search engine returned (10-point Likert scales). In another open-ended question, they were asked to provide a detailed explanation of the search strategy used.

### **3.5.1.2 PARTICIPANTS**

The twenty-four participants were students and staff from the University of Bath, who were recruited via an online advertisement and posters on notice boards, 12 males and 12 females. No participant had taken part in the earlier studies. The participants were aged between 21 and 41, with the average being 29.7 years. All reported that they used search engines in their everyday lives. They were each given five pounds in cash in gratitude for donating their time. In addition, they were motivated to search for good quality web pages by competing for two cash prizes of 30 pounds.

### **3.5.1.3 MATERIALS**

The version of the HCI browser used in Experiment 2 was altered only by the addition of a digital clock in the top right-hand corner, which counted down the time remaining for each task. Figure 3.10 shows an example of search interface presenting 24 options in the Random ranking condition with a digital clock.



**Figure 3.10 Experiment 3: An example of a computer screen presenting twenty-four options (on four SERPs) in the Random ranking condition with a digital clock.**

### 3.5.1.4 PROCEDURE

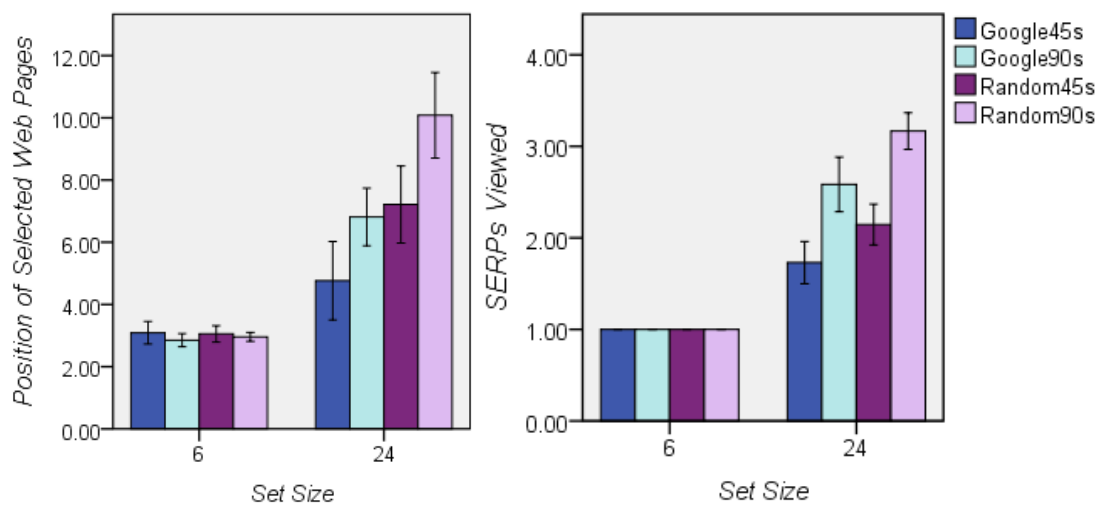
The procedure closely followed that for Experiment 1 and Experiment 2. The participants had two training tasks to complete, so as to experience both time conditions. The questions for the training were “Why did ancient Egyptians mummify their dead?” and “How does economics affect our daily life?” within time limitations of 45 seconds and 90 seconds, respectively. Pre-specified search terms, which were abbreviated versions of the problem specification, were used and indicated to the participants. For example, when the task question was “Why is meditation sometimes recommended for managing stress?”, the search term was “meditation for managing stress”. They undertook the training tasks with the same interface and time constraint as would apply for their experimental session. The browser window displayed a task question, a search term, a ranking type and a timer at the top of the window. This information remained throughout the search task and the subsequent ratings. During each, a beep sounded and a visual sign appeared, if and when ten seconds remained. In a debriefing session, the purpose of the experiment was explained to the participants, and they were allowed to ask any question related to the study.

### 3.5.2 RESULTS

The main analyses reported below are 2x2x2, Set Size (6 vs. 24) x Ranking Type (Google vs. Random) x Time Limitation (45 seconds vs. 90 seconds) mixed ANOVAs.

#### 3.5.2.1 POSITION OF THE SELECTED WEB PAGES

Figure 3.11 (left) shows the mean ordinal position of the selected web pages. There were significant main effects of Set Size ( $F(1,22) = 20.3, p = 0.000, \eta^2_p = 0.48$ ), Ranking Type ( $F(1,22) = 13.5, p = 0.001, \eta^2_p = 0.38$ ), and Time Limitation ( $F(1,22) = 5.14, p = 0.034, \eta^2_p = .19$ ). The interaction effect between Time Limitation and Set Size on the position of the selected web pages was significant ( $F(1,22) = 6.7, p = 0.017, \eta^2_p = 0.23$ ). The interaction effect between Ranking Type and Set Size on the position of the selected web pages was also significant ( $F(1,22) = 12.9, p = 0.002, \eta^2_p = 0.37$ ). Those in the large set condition selected web pages from further down the list than did participants in the small set condition, especially when they selected from the Random-ordered SERPs and when they had 90 seconds available to make their selection.

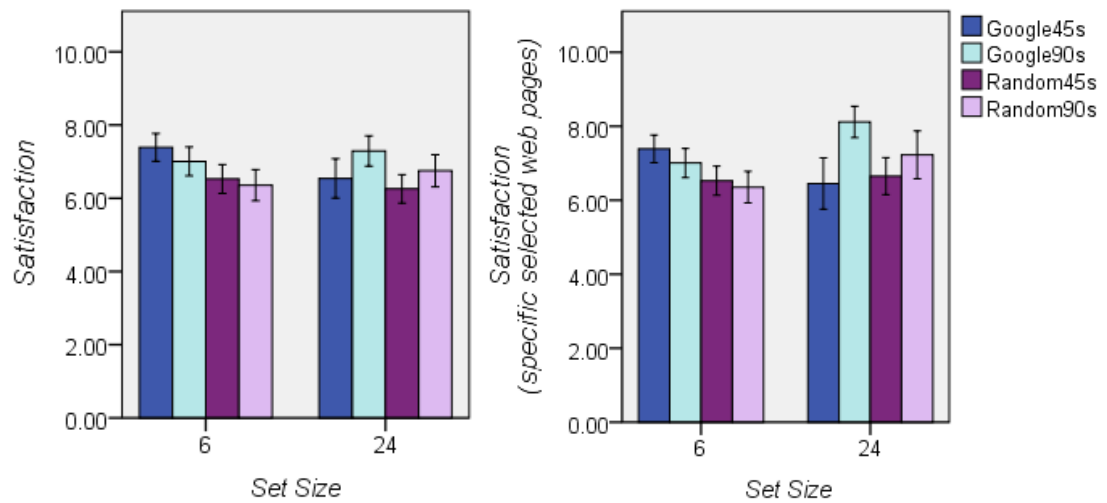


**Figure 3.11 Experiment 3: Mean position of the selected web pages (left), and number of SERPs viewed before selecting (right). Error bars show standard errors.**

According to a one-sample t-test, the number of SERPs viewed in the large Set Size was significantly greater than 1 ( $t(11)=12.76, p < 0.001$ ). The ANOVAs revealed that the main effects of the independent variables on the number of SERPs viewed in the large set condition were significant, Ranking Type ( $F(1,22) = 6.77, p = 0.025, \eta^2_p = 0.38$ ), and Time Limitation ( $F(1,22) = 25.34, p < 0.001, \eta^2_p = 0.68$ ). There was no significant

interaction effect between Ranking Type and Time Limitation ( $F < 1$ ). Participants in the large set condition browsed through more SERPs when they had to select a web page from the Random results list. In the time limitation of 90 seconds, those in the large set browsed significantly more SERPs than they did when they had 45 seconds to select a web page.

### 3.5.2.2 RATING OF THE SELECTED WEB PAGES



**Figure 3.12 Experiment 3: Mean satisfaction with the selected web pages (left), and mean satisfaction with selected web pages in which the data were pruned (right).**

**Error bars show standard errors.**

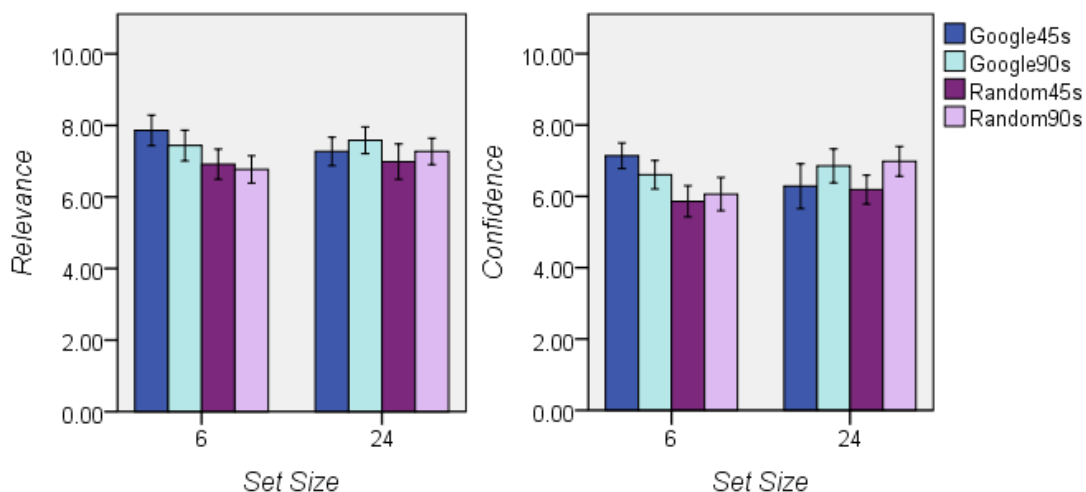
Figure 3.12, left, shows judged satisfaction with the chosen web pages in each cell of the experiment. The main effect of Ranking Type on the participants' mean satisfaction was significant ( $F(1,22) = 6.31, p = 0.02, \eta^2_p = 0.22$ ) and they were more satisfied with pages from Google-ordered lists than from Random-ordered ones. There was no significant main effect of either Time Limitation or Set Size on their mean satisfaction ( $F(1,22) = 0.75, p = 0.39$  and  $F(1,22) = 0.05, p = 0.81$ , respectively). However, the interaction effect between Time Limitation and Set Size was significant ( $F(1,22) = 5.13, p = 0.034, \eta^2_p = 0.19$ ). In both the Google and the Random conditions, satisfaction in small set sizes relative to large sets increased as time pressure increased.

The data were pruned so that only those chosen web pages present in the small set were considered in all cells of the experiment (otherwise, reduced satisfaction in selections from the large set could be a peculiarity of the actual web pages that were chosen).

Every participant contributed at least one such judgement in all large Set Size conditions (Figure 3.12, right). There was now no significant main effect of Ranking Type on the participants' mean satisfaction ( $F(1,20) = 3.42, p = 0.079$ ). Nevertheless, importantly, the significant interaction effect between Set Size and Time Limitation on their mean satisfaction ( $F(1,20) = 10.04, p = 0.005, \eta^2_p = 0.33$ ) was maintained.

Ranking Type, Time Limitation and Set Size did not have significant effects on Trust judgements. There was no significant interaction effect, with all  $F$ s being close to 1 (Trust was significantly correlated with Satisfaction, with  $r = 0.60, p < 0.001$ ). In addition, Relevance was strongly correlated to Satisfaction ( $r = 0.78, p < 0.001$ ) and the main effect of Ranking Type on Relevance was significantly different ( $F(1,22) = 6.18, p = 0.021, \eta^2_p = 0.22$ ). The participants thought that their selected web pages in the Google-ordered lists were more relevant to the task questions than those in the Random-ordered ones (Figure 3.13, left). There was no main effect of Time Limitation and Set Size and no interaction effect on Relevance (all  $F$ s  $< 1$ ).

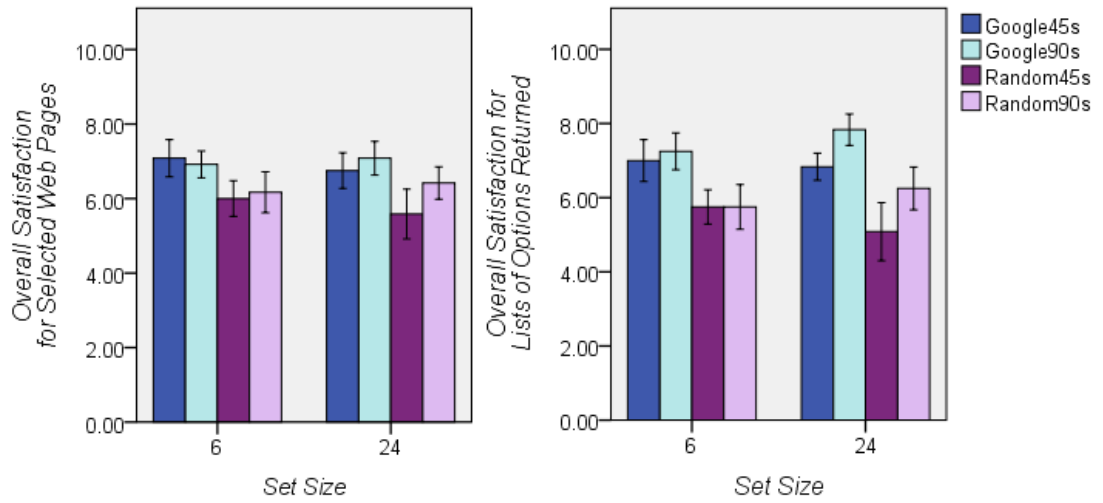
There was no significant main effect or interaction effect on Confidence judgements, with all  $F$ s being close to 1. However, Confidence was significantly correlated with Satisfaction ( $r = 0.86, p < 0.001$ ). This indicated that if the participant was satisfied with the selected option he/she was more likely to be confident that the selected option was the best in the set available.



**Figure 3.13 Experiment 3: Mean relevance (left), and confidence (right).**

**Error bars show standard errors.**

### 3.5.2.3 OVERALL RATING OF THE SELECTED WEB PAGES



**Figure 3.14 Experiment 3: Mean overall satisfaction for the selected web pages (left), and mean overall satisfaction for the lists of options returned (right). Error bars show standard errors.**

Figure 3.14 shows the participants' satisfaction ratings at the end of each task-block (i.e. after each ranking type/time limit). A main effect of Ranking Type on overall satisfaction for the selected pages was significant ( $F(1,22) = 11.42, p = 0.003, \eta^2_p = 0.34$ ). In both large and small set sizes, the participants were significantly more satisfied with their selected web pages from the Google-ordered lists than with the Random-ordered. However, the main effect of Set Size was not reliable ( $F(1,22) = 0.032, p = 0.86$ ). There was no significant interaction effect. A main effect of Ranking Type on overall satisfaction for the lists of options returned by the search engine was significant ( $F(1,22) = 23.38, p < 0.000, \eta^2_p = 0.52$ ). The participants were satisfied with the result lists ranked by Google more than those ranked at random. There was no significant main effect of Set Size or interaction (both  $F$ s  $< 1$ ).

### 3.5.2.4 STRATEGIES FOR CHOOSING A WEB PAGE

The analysis of the qualitative data involved the same approach as for the previous experiments. This resulted in five main themes: *keyword matching*; *content comparison*; *ranking of options*; *recognised information sources*; and *reliability of information sources*. The findings provided insights into the participants' searching strategies, which in part were influenced by the experimental conditions. Although the search strategies



were categorised into themes, most of them used a combination in order to choose a web page. For example:

[Participant #14, Large set, Random ranking, 45 seconds] *“Read the address of the web page first, then look at the title. If can’t decide, use the abstract.”*

First, *keyword matching* pertained to statements where participants mentioned searching through the titles and snippets for keywords related to the task question:

[Participant #2, Large set, Google ranking, 90 seconds] *“I then focus on web pages with keywords relating to the questions asked.”*

[Participant #8, Large set, Random ranking, 45 seconds] *“Headlines that were related to the task question.”*

[Participant #15, Small set, Google ranking, 90 seconds] *“Catch my eye on the boldface keywords.”*

Second, *content comparison* referred to statements where relevance judgement was mainly made based on the content in the snippet. Some participants commented that they used snippet content to judge the relevance of web pages in the Random ranking condition since the order of results were not reliable. For example:

[Participant #1, Small set, Random ranking, 45 seconds] *“A heavy reliance of the information in the blurb and URL. Without the page rank more time has to be spent reading the information and since position has no bearing on relevance, each link must be studied to determine its usefulness.”*

[Participant #6, Large set, Google ranking, 45 seconds] *“Try to evaluate the websites by carefully reading the preview on the search engine results.”*

[Participant #11, Small set, Google ranking, 45 seconds] *“To read all the summaries of the different web pages and chose the one which seemed to be the most specific to the question.”*

Third, *ranking of options* pertained to the participants’ selection strategies that were based on their perception of the reliability of the ranking algorithms. The participants mostly used this strategy when they selected web pages from the Google-ranked lists and when time was more strongly a constraint. For example:

---

[Participant #4, Large set, Google ranking, 45 seconds] *“Usually choose the first few choices available from the page 1.”*

[Participant #12, Large set, Google ranking, 45 seconds] *“Usually, Google gives quite good advice on choosing the link, so the first few links will be related to my task.”*

Next, *recognised information sources* referred to statements mentioning recognised or familiar web sites. It is worth noting that participants typically reported the use of this strategy when they were under time pressure. For example:

[Participant #1, Small set, Google ranking, 45 seconds] *“Scanning the blurb provided for each page in combination with an assessment of the author and website. For example, an xyz.about.com article is likely (In my personal experience) to be less useful than a Wikipedia entry.”*

[Participant #2, Large set, Random ranking, 45 seconds] *“First I look at the link to see if it’s an educational or institutional authority.”*

[Participant #18, Large set, Google ranking, 45 seconds] *“First of all, looking at the website name and trying to think whether the website is familiar.”*

Finally, *reliability of information sources* pertained to a decision made based on the trustworthiness or authorisation of the web sites:

[Participant #6, Large set, Random ranking, 45 seconds] *“Tended to choose from well-known and trustworthy websites instead of random ones. I believe that the source of the information is far more important than the information itself.”*

[Participant #11, Small set, Random ranking, 45 seconds] *“Select the one which seemed to answer the question best and be a trustworthy source.”*

[Participant #17, Small set, Google ranking, 90 seconds] *“The results’ summaries were less helpful on these questions, so I mainly went by the reputation of the website’s creator.”*

The qualitative findings provide insights regarding search strategies and suggest that the participants changed their strategy of selecting web pages according to the perceived reliability of the ranking algorithm and the time constraint.

### 3.5.3 DISCUSSION

These experiment outcomes indicate that time pressure played a crucial role in determining the effect of set size on satisfaction with the results of the choice process. For, when time was more strongly constrained, choices from small sets led to relatively more satisfaction. The ranking type, set size of options, and time pressure did not significantly affect participants' confidence. However the confidence of the decision made was significantly correlated with their satisfaction.

The main effect of ranking type on participants' satisfaction was significant, but its interactions with the other independent variables were non-significant. In addition, the participants were more satisfied with the selected web pages when they chose from the Google-ordered lists than from the Random-ordered lists regardless of time limitation or the number of options provided. It is worth noting that this may simply be because the quality of the web pages selected from the Google-ordered lists was trusted more by the participants.

### 3.6 GENERAL DISCUSSION

The series of experiments presented in this chapter sought to explore the effect of choice overload and its moderators in search engine use. More specifically, the potential influences that impact on satisfaction judgement regarding the selected option were investigated.

The tasks used were open-ended problems on various topics. It seemed crucial to choose tasks that information sources could be considered to satisfy *to various degrees* (unlike fact-finding tasks), so as to allow sensible and varied satisfaction judgments. In addition, the use of such tasks helped to ensure that the participants were given tasks that they were not familiar with, which encourages the possibility of choice overload (which is known to be less of an issue for familiar choices, Mogilner et al., 2008). Of course it is true that not all information seeking tasks are of this kind, but many are, and the interest in these experiments was to explore tasks for which the phenomenon of choice overload might be revealed.

In contrast to Oulasvirta et al. (2009), the results have shown that there is no evidence that choice overload will typically affect satisfaction with web pages found through

---

keyword searches. Apart from non-significant results, in most conditions of the current research the tendency was for larger numbers of returns to be associated with better subjective outcomes regarding judgements of individual web pages in Experiment 1 and for overall satisfaction in Experiment 2, which contrasts with these authors' outcomes, but only marginally.

It was predicted that the ordering of search engine returns would mitigate choice overload effects and thus, unsurprisingly it emerged that in an ordered environment, the high expectation of gaining a better result may not be triggered by the large number of options, but rather, the perceived effectiveness of a ranking algorithm. Furthermore, the outcomes suggest that the regret about not consuming unchosen options may be diminished when the options are ordered, such that choosing from the top of the list could ensure their quality to some extent. In addition, the participants may treat reliability of ordering as reducing uncertainty about whether high-value options are likely to be located further down the list of search returns. That is, under the perceived reliable ranking condition, the participants may have set a threshold value of a targeted option based on the value of items presented earlier on the result list. Hence, the values of options located deeper through an ordered list would have been predicted to be likely to be worse than the threshold and seem less beneficial to consider, which in turn makes these options less likely to be chosen. However, when the ordering of options was perceived to be random, the threshold value may have been sensitive to context, that is, it could have been adjusted based on the options assessed up to that point during search process (as in the models of Cox & Young, 2004; Brumby & Howes, 2008) and the selection was likely to be made through this estimation of the relevance of the options' content. The results from the second experiment confirmed that ordering does indeed affect selection processes, in that, when the participants believed the results had been ordered by experts or novices they were more likely to choose a web page from earlier in the list, compared to random ordering, even when the order of items was identical. Similarly, in the third experiment the participants chose earlier items from Google-ranked lists than from randomly ordered ones. These ordering effects confirm and extend previous work, which has shown that Google users typically select from the first page of returns (Pan et al., 2007) and that the order of a document set affects judgements of relevance (Purgalis et al., 1990). However, it is worth noting (by comparing the number of SERPs viewed with the position of the selected web pages) that the participants often

did not select a high-value web page when it was first found, but browsed more options through a results list on other SERPs, before returning to the first SERP to select the option even when ordering was perceived as reliable. This suggests that the participants may seek to increase their confidence on their decision by assessing a certain number of options to ensure that there was no better choice presented later on the lists.

Furthermore, these effects of ordering do not seem to affect choice overload, i.e. the effects of set size on subjective judgements did not interact with beliefs about order or the actual order of search returns. Instead, evidence has been found that time pressure is an important determinant of choice overload. For instance, in Experiment 3, the relative satisfaction with the web pages chosen from the larger as compared with the smaller sets interacted significantly with the time available to choose. The role of time pressure from the experiments is also important for theoretical reasons, because it supports the “process-product leakage” account of choice overload, rather than regret or prior expectation. Finally, the work reported on in this chapter confirms that interface design can affect process and strategy in such a way that user satisfaction with retrieved information is affected.

### **3.6.1 QUALITATIVE FINDINGS**

The results from the qualitative responses suggest that participants generally used a combination of search strategies to achieve their goal. Although the sets of themes identified from the three experiments are slightly different to some extent, they commonly represent fundamental strategies used in information seeking activities.

The sequence of strategies employed varied based on participants’ preference and the perceived value of each strategy. Some participants mainly focused on the available information of the titles and snippets of the search results and changed their strategy when having difficulty in differentiating the level of relevance from that information. The use of strategy was also influenced by the time pressure, for example, when participants had less time, they chose not to verify all available information because only a few options could be examined. In such cases, the strategy that was reliable and consumed less time appeared to be more suitable. For instance, they tended to rely on their simple recognition and based their decision on the familiar web sites. Furthermore, some participants relied on the order of search results when they regarded the ranking algorithm as being reliable, whilst others mentioned having difficulty in selecting

---

options when they decided not to depend on the ranking algorithm as they did not trust its efficiency. Despite providing insights into search behaviour that support the results from the quantitative analysis, the qualitative findings reveal that the selection behaviour was mostly guided by the estimated relevance of information provided by web pages and the reliability of information sources. In addition, the estimated relevance of an item was not based purely on the content provided but also influenced by beliefs and perceptions on reliability of ranking algorithms.

### **3.6.2 LIMITATIONS AND FUTURE WORK**

The findings challenge search engine designers to be sensitive to time pressure. It could be argued that the way time pressure affects users' needs and satisfaction is likely to vary across search contexts. However, in the case that users need to find information rapidly, receiving a large set of returns may affect their decision satisfaction and behaviour to some extent. Further investigation is needed aimed at providing greater robustness in support of the outcomes from such a specific case.

One important limitation of these experiments is that the participants were asked to choose single web pages. In many contexts, multiple information sources will be sought and under such circumstances a variety of search returns will be required. Consequently, the search results should be diversified such that results from different categories could be presented early on the list, possibly in a hierarchical structure. Indeed, such a design might be one way in which search interfaces could be sensitive to time pressure whereby users could consider in turn each specific category where the number of options was considerably fewer than the entire results list. In addition, with this configuration and less time pressure, a user could browse through all the categories to explore and make comparisons between options.

## **3.7 CHAPTER SUMMARY**

In this chapter, the methods and results of a series of three experiments have been reported. In the first, the participants' satisfaction levels were examined by providing two different set sizes of options (six options and twenty-four options), with half being allowed to perform iteration (changing their search term). The second experiment continued to explore the effect of different set sizes of options on participants'

satisfaction. In addition, the effect of the perceived orderings of search results was investigated, with sets of options being generated by Google with pre-defined search terms, which were used across all different ranking conditions. In the third experiment in addition to different set size options and ranking algorithms, different time limitations were used to investigate the boundary of the choice overload effect.

Through these three experiments it has been found that:

- The perceived reliability of the ordering of options affects search and selection behaviours. That is, users only browsed through a few SERPs when the large set of options was available and mostly chose the options from near the top of the lists when the ordering was perceived as reliable.
- Satisfaction judgement regarding the selected options is influenced by the ordering of search returns such that the more reliable the users perceived the ordering, the more satisfied they were with their selected options.
- Large numbers of search results returned from a search engine tend to be associated with more satisfaction on the selected options.
- The satisfaction on selected options is affected by the number of available options interacted with the time pressure. For, when time is more strongly constrained, choices from small sets led to relatively more satisfaction.
- Time pressure is the important factor that moderates the choice overload effect, which supports process-product leakage theory.

In the next chapter of this thesis the choice overload effect with a larger choice set is explored, with the aim being to assess the possibility of differentiating between the satisfaction judgement regarding the selected option and that in relation to the search and selection process. Furthermore, the variety seeking and the chance encountering in search engine use are probed by including a short list of suggested supplementary web pages as extra options.

## CHAPTER 4

# VARIETY SEEKING AND CHANCE ENCOUNTERING

### 4.1 CHAPTER OVERVIEW

In Chapter 3, no significant main effect of set size was found on satisfaction with selected web pages. In particular there was no hint of a choice overload effect; this contradicted the choice overload phenomenon found in some previous studies (Iyengar & Lepper, 2000), including the only known study using a search engine task (Oulasvirta et al., 2009). However, these findings were not so unusual based on the meta-review of fifty experiments that manipulated the number of alternatives (Scheibehenne et al., 2010).

The work presented in this thesis has so far concentrated on investigating the effect of choice overload by manipulating a number of options, with the improved experimental design, so as to be consistent with previous research in the search engine use context. However, it must be conceded that 24 items is a very limited model of a “large” number of search engine returns. This chapter continues to address **RQ1: Does choice overload affect users’ satisfaction with search engine use?**, by expanding the number of options in the largest set, i.e. the aim is to investigate the effect of choice overload with a larger number of options.

Furthermore, the experiments presented in Chapter 3 has demonstrated that the participants, even in the larger option set condition, usually selected a web page from the first results page, often without bothering to iterate their search terms or browse options



on later pages. This reliance on the ordering of options may limit the chance of encountering options representing different aspects of the field of interest. It is quite possible that options with high quality but not widely known may be ranked towards the end of the results list by the obscure ranking algorithm, which may reduce the chance that users will come across them.

The third research question is introduced in this chapter, with its underlying purpose being to investigate the effect of variety of search results on users' behaviour and serendipitous experiences – **RQ3: Does a search task which encourages users to seek various targets affect users' behaviour and satisfaction in search engine use?** This question is addressed by asking participants to collect three web pages relevant to the same topic. This simple extension of the task will, it is presumed, motivate variety seeking, and consequently affect the utility of the search engine's ranking of returns.

The results from the experiment reported in this chapter reveal that a large set of search returns yields more satisfaction; even in the situation that time is a constraint. The findings also suggest that in a single choice situation users are more likely to trust and rely on the ranking algorithm to providing valuable options, which may reduce the chance of discovering new quality options. Also, in a multiple targets selection task the variety of options leads to more satisfaction with the sets of selected options and with the search and selection process. In addition, the search and selection process is less affected by the ordering of search results when variety is required.

---

## 4.2 INTRODUCTION

The findings from the experiment by Oulasvirta et al. (2009) showed that choosing from the large set of search results yielded relatively less satisfaction regarding the selected options than choosing from the small one. In contrast, the previous work in this thesis (See Chapter 3), revealed that a large set of options could lead to more satisfaction with selected web pages, despite there being a time constraint. The experiments presented in Chapter 3 were developed from Oulasvirta et al.'s work, but used an online environment that more closely modelled real search engine use. Nevertheless, it could be argued that the larger set of options manipulated in Experiments 1, 2 and 3 was simply not big enough to create an overload effect. Perhaps users who use search engines in their daily lives are accustomed to the typical extensive set of returns from search engines, and this has improved their cognitive ability or search strategy to deal with more search results in the online environments (Bilal, 2000; Malhotra et al., 1982). Therefore, the new experiment reported in this chapter aims to investigate further the effect of choice overload in search engine use by examining the users' satisfaction with the selected options being chosen from a larger set of options than in Chapter 3.

Another reason why the large options set tended to produce a positive effect in Experiments 1-3, may be because the search results were ordered, whereby more valuable ones were likely to be presented near the top of the list. Users may perceive that having more options to choose from increases the chance of finding more preferred results, given that they can fall back on relying on the ranking of options when they have difficulty in making a decision. This is to some extent supported by the previous findings, which showed that the ordering of returns affected the search behaviour and satisfaction on the selected options. More specifically, the work reported in Chapter 3 demonstrated that the participants were more likely to select web pages from the first SERP and were relatively more satisfied with their selected pages when their options were ranked by a perceived to be reliable ranking algorithm. Additionally, they were less likely to iterate or modify their search terms when they were selecting from a large set of options. This behaviour appears to be common among search engine users, as described in a comparison of nine search engine transaction logs (Jansen & Spink, 2006). In that study, it was found that the percentage of searchers viewing only the first SERP was exceptionally high and significantly increased over the years, 1997-2002, across both US

and European-based Web search engines and moreover, searchers mostly searched by using only one query (Jansen et al., 2000).

As already noted, in a real world online search environment, search results are typically ordered according to some ranking algorithm, which gives the perception that the most valuable options are likely to be presented early in the results list. The ordering of returns yields positive effects on the search process in several ways, given that the users are likely to rely on the ranking algorithm. For example, in simple fact-finding tasks obviously users spend less time in the search and selection process by selecting the options presented at the top of the list. Furthermore, in open-ended problem tasks the ordering of returns seems to help searchers find relevant search results from a long list of search returns to some extent. Although it seems that the ordering of options provides a benefit for search engine users as it mitigates the effect of choice overload, relying on it can reduce the chances of encountering quality options in the following ways. Firstly, it is of course possible that the highest quality web pages may be located elsewhere toward the end of the results list. Those web pages may be recently published or less popular, which seems to be judged as being of lower quality by some ranking algorithms (e.g. Page et al., 1999). Consequently, unknown web pages with high quality can be easily missed and may not be able to be found by users if they are late in an ordered list. Furthermore, the use of a single search query can limit the boundary of the information gathering due to the fact that the topic of interest may be represented by different terms. By using only one search term, users may have a poor chance of knowing what options they are not considering, especially regarding the different aspects entailed in the subject of interest.

Users' potential over-reliance on ranking algorithms has motivated some researchers to explore chance encountering and serendipity in information seeking contexts. Aiming to reduce the bias from popularity-based ranking, Pandey et al. (2005) conducted an experiment that introduced randomness into a search result ranking algorithm. A thousand links of joke or quotation pages were presented on the web site in descending order of funniness rated by the visitors to the web site. Over a period of 45 days, each was able to rate whether the joke/quotation pages were funny, neutral or not funny and the items that were ranked near the top of the list indicated high popularity. In the condition that rank promotion was used, items that had not been seen and rated by any visitors were randomly inserted on the list starting at ranking position 21. The ratio of

---

funny votes to total votes was measured for two conditions, in one rank promotion was used and in the other it was not. The results revealed that the ratio obtained using rank promotion was larger than that achieved using solely popularity ranking. This showed evidence of a popularity-based ranking bias and demonstrated the advantage of partially randomising rank positions in order to promote unknown-quality options.

Further, Toms and McCay-Peet (2009) investigated chance encountering in information retrieval on a WikiSearch-integrated system. A list of five suggested Wikipages was generated based on the currently selected page from the results list, with these pages being generated by using the first paragraph of the currently selected page as a search string. The participants were asked to respond to a problem task by adding relevant Wikipages to a collection and to answer open-ended questions at the end of the experiment. The use and perception of the suggested pages were assessed and the results revealed both advantages and drawbacks. Whereas some participants reported that the suggested pages helped to broaden or narrow their search and helped to generate ideas for further searches, others stated that the suggested pages distracted them from the tasks at hand. The authors proposed that individual differences could lead to different levels of openness to experience, and that a design to enhance chance encountering could be viewed as a distraction, particularly during a goal-directed task.

Concerning the issue that authors and searchers are likely to express the same thing in different ways, Kristensen (1993) suggested that a tool connecting the different names of a similar concept is the key to increasing the coverage of related search results. In this author's query expansion experiment, each query was a search in five distinct modes: basic search with the original query, synonym search, narrower term search, related term search, and the union of all previous searches. For the last four modes, the search term was extended from the original query by adding extra terms disjunctively, for example, for the synonym search mode the pattern of search term was "original query + or + synonym". Although the results from each search mode could contain those from the other modes for the same original query, the results showed that different documents were retrieved when searching by using a different inclusion of terms. The findings revealed that the two search modes, namely, the synonym and related term, retrieved most relevant articles.

The work presented in this chapter aims to investigate the advantages of coming across the different sets of search results in two ways: by promoting some web pages from further down the results list, and by presenting some results from searching by using synonym search terms. These two ideas each lead to a relatively simple, but novel design, in which “supplementary” returns are presented alongside the conventional list of search returns.

At the same time, this experiment introduces a more discriminating set of satisfaction judgements, to try to throw extra light on the process-product leakage theory. According to this theory satisfaction with the process can leak into the judgement about the selected option. The experiment presented in this chapter is designed to investigate the relation between satisfaction with the selected option and with the search and selection process, by simply asking participants to make separate satisfaction judgements for each of these (it is supposed that the presence of a “satisfaction with process” question will in turn shift the interpretation of “satisfaction with selected option” to a purer judgement). Moreover, the satisfaction on the selected option is measured according to two aspects. The first is that the selected option is judged based on its value to the current task, whilst the second pertains to the judgement being made independently of the task. (It seems plausible that these two aspects might be merged in the simpler question, and the new questions allow this possibility that they behave separately to be explored).

A further innovation of the experiment in this chapter is that it introduces a new task type, in which multiple webpages on the same topic must be found. It seems plausible that in such tasks, the user may seek to increase variety among those selected items. A list of search results ordered by a single ranking algorithm may not serve this purpose very well, which may lead to negative effects on the decision process and the selected option. In addition, it is expected that a search task which encourages users to choose multiple options will affect their search behaviour such that they may seek variety for a set of selected options.

Variety seeking behaviour has been investigated in many choice contexts mostly with an unordered choice set with the aim being to explain the strategy used in the seeking process. Ratner, Kahn and Kahneman (1999) conducted a series of experiments to examine whether individuals tended to seek variety by switching to choosing less-preferred items rather than repeating their choice for more-preferred ones. The

---

enjoyment from switching to less-preferred items was measured through choosing songs and the results revealed that the participants did choose less-preferred ones when they were provided with a high-variety set. However, this resulted in less enjoyment for the selected songs in this condition. Conversely, when they were asked about their overall experiences after completing the tasks, their ratings revealed that high-variety sets were viewed more favourably than low-variety ones. That is, this experiment showed that somehow they enjoyed the selection process more when more variety was provided.

Another behaviour with regards to seeking for variety was reported by Fox, Ratner and Lieb (2005). Their series of studies demonstrated that consumers tended to have a bias for even allocation when choosing products or services; allocating money to beneficiaries, purchasing products for future time periods, and when choosing from a menu of consumption options. In addition, they found that even allocation was mostly based on how the available options were grouped either by individual preferences or the pre-defined categories. This phenomenon they termed “partition dependence”, referring to the tendency to make different allocations among the same set of options based on the way those options are subjectively grouped. In the wine selecting study, the participants were asked to select three wines from a list of six white wines that were in the same price range and had the same range of quality scores. The participants were also grouped into levels of expertise, novices and experts, based on the number of bottles of wine they had purchased in the previous year. The results revealed that the participants were more likely to choose three different kinds of grape when wines were grouped by type of grape, and they were more likely to choose wines from the three different regions when the options were grouped by this aspect. Moreover, the participants who were categorised as novices were more affected by partition dependence than the experts. In particular, this study showed that people with strong preferences will rely less on diversification heuristics and that they are less likely to be affected by the partitioning of options. Further, even allocation behaviour in another context was reported by Benartzi and Thaler (2001), who demonstrated that some investors divided their contributions evenly across the funds offered in a savings plan.

Returning to variety seeking in search engine use, on the basis of the literature, it would appear that search engine users can adapt their search strategy in several ways in order to increase the variety among the selected set of returns. Firstly, they may select less-preferred options because of the relatively positive experience of the search and selection

process, even though this may lead to less satisfaction when compared to the individual choice. Secondly, they may be influenced by even allocation bias when trying to increase variety among the selected options. In addition, they may search deeper and select an option that is located further down the results list, especially when they are provided with a large set of search results. Finally, it seems that users may choose unfamiliar or unknown options included in the selected options set as an exploration or for enlightenment. All these strategies seem plausible when seeking variety from the set of options returned by a search engine. However, a list of search results that is ordered by one ranking algorithm may complicate the seeking process, such that it could be difficult for users to search for options with different features from the list.

In general, it seems that variety seeking behaviour is an important and underexplored issue in the search engine use context, especially in a situation where several information sources and a range of contents are required. That is, when users search deeper through an ordered list aiming to seek for variety, an extensive set of returns with a slight difference between each item may lead to less satisfaction with the decision process and the selected options. However, it has been suggested that the perception that differences among options already exist might reduce variety seeking behaviour (Van et al., 1996). It is expected here that introducing variety among search results may increase satisfaction with the search and selection process. The lack of attempts in prior research to investigate variety seeking behaviour and its consequences in an ordered environment, such as search engine use, provides the motivation to explore this and thus develop a better understanding on user behaviour in this context.

### **4.3 EXPERIMENT 4: NUMBER OF OPTIONS AND SUPPLEMENTARY WEB PAGES**

#### **4.3.1 METHOD**

##### **4.3.1.1 DESIGN**

This experiment was designed to extend the investigation on choice overload in search engine use and to explore variety seeking behaviour and chance encounters by providing additional suggested web pages. Firstly, a larger set of options returned by Google was introduced: forty-eight in the large set size condition. The Set Size of a results list was a between-subjects independent variable, with three levels: There were six items (on a

---

single SERP), 24 items (on four separate SERPs) or 48 items (on eight separate SERPs, each reached by pressing a page number button at the bottom of each SERP). The second independent variable was the Selected Set, that is, half of the participants were asked to select and open one web page per task, whereas the other half were asked to select three web pages for each task.

Next, an additional search results list was manipulated; a list of two supplementary web pages was added on top of the main Google result list. The participants were informed that the supplementary web pages were generated in one of two conditions:

- *Synonym search*: the first noun in a search term was substituted by its synonym. The synonym search term was then used in Google to retrieve a supplementary web page (a supplementary web page was selected from the top of a synonym-search results list). Two different synonyms were used to generate two supplementary web pages.
- *Quartile search*: the first web page located in the third quartile and fourth of the same search results list were used as supplementary web pages. Given the search terms were pre-defined, it was possible to select these manually prior to conducting the experiment as the generated results list would always be the same.

The synonym search returns were generated as follows. A simple stemming algorithm was developed to identify automatically the first noun in the search term. Two synonyms were manually selected from the common synonyms found across four different thesauri: WordNet thesaurus (About WordNet, 2010; Miller, 1995), Oxford thesaurus (Oxforddictionaries.com, 2013) Roget's 21<sup>st</sup> century thesaurus, (Thesaurus.com, 2013), and Collins thesaurus (Collinsdictionary.com, 2013). To make this manipulation plausible pre-specified search terms were used and indicated to the participants, which were abbreviated versions of the problem specification. For example, when the task question was "How can we protect our privacy online?" the search term was "how to protect privacy online", and the first and the second synonym search terms were "how to protect seclusion online" and "how to protect confidentiality online". Hence, the main results list and the supplementary list presented in each condition for each task would always be the same.

Types of supplementary web pages were manipulated as a within-subject independent variable and an additional condition with no such web pages was included as a control.



The tasks were divided into three sessions (four tasks per session) based on the supplementary type, i.e. each session used one of: no supplementary, synonym supplementary or quartile supplementary. The supplementary types were ordered such that each appeared in ordinal positions equally across the participants. For each of the twelve tasks, under the single choice condition the participants were given 45 seconds to select a single web page that provided valuable information from the list of search results. If they opened a selected web page, they were not allowed to reject it and search for another, which meant their decision had to be made entirely from the snippets. After each web page was chosen, they inspected the page and then rated their satisfaction with it, the list of search results, and the search process (using the satisfaction questions listed below). They were also asked to rate their confidence that their selected web page was the best option in the search results list. In the multiple targets condition, i.e. selecting three web pages per task, the participants were required to select and open exactly three web pages per task. They were given 45 seconds to choose each single web page, followed by no time limitation to consult the content of the selected web page before selecting the next.

### **4.2.1.2 PARTICIPANTS**

The forty-eight participants were undergraduate and post-graduate students from the University of Bath, 24 males and 24 females. They were recruited via an online advertisement and posters on notice boards. The participants were aged between 18 and 41, with the average being 27.3 years. All of them used search engines in their everyday lives. They were given five pounds in cash as appreciation for donating their time and they were motivated to search for good quality web pages, as the best of these were to be awarded two cash prizes of thirty pounds.

### **4.3.1.3 MATERIALS**

The HCI Browser (Capra, 2010), an open-source extension to Mozilla Firefox 3, was modified to collect data and to guide the participants through tasks. The SERPs were altered using Google's API and all interactions were logged. After they had selected and consulted the content of the web page, the participants were asked to rate their satisfaction with that chosen. In addition, the ratings were designed to capture their satisfaction regarding two aspects: the content of the selected web pages and the search and selection process.

---

All ratings were made on 10-point Likert scales and the instrument for this was as follows:

For the single choice condition,

- You are satisfied with your selected page with respect to the information it provides concerning your question.
- Your selected page was satisfying in content, independent of your question.
- The total list of [number of options] web pages returned by the search was satisfactory.
- The search and selection process was satisfactory.
- I am confident that I chose the best web page from the available set.

For the multiple choices condition,

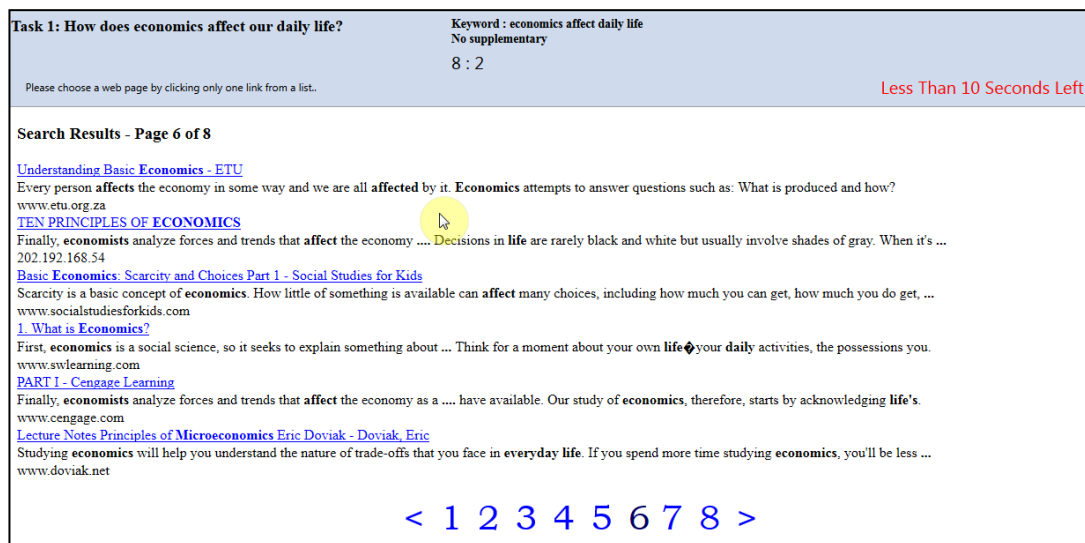
- You are satisfied with your three selected pages with respect to the information they provide concerning your question.
- Your three selected pages were satisfying in content, independent of your question.
- The total list of [number of options] web pages returned by the search was satisfactory.
- The search and selection process was satisfactory.
- I am confident that I chose the best three web pages from the available set.

After each single session was completed, a questionnaire asked the participants about their overall satisfaction with their selected pages and that for the list of pages that the search engine returned (10-point Likert scales). Additionally, the open-ended questions asked them about the strategy they used to choose web pages and the usefulness of the supplementary web pages in the novel designs.

#### **4.3.1.4 PROCEDURE**

The participants completed the experiment individually in a laboratory, with the experiment for each participant being divided into a training session and an experimental session. On arrival at the laboratory, participants were introduced to the HCI Browser interface and instructed that their general task would be to search for web pages that provided valuable information for a series of separate questions or tasks. Two types of

supplementary web pages were described to them in written instructions and questions were answered by the experimenter. Then, the participants completed two sample tasks in the training session, so as to experience the search results list with supplementary web pages and without them. The questions for the training were “How does economics affect our daily life?” and “How can we manage time effectively?”, with the experimenter available to offer guidance as required. After each participant confirmed that they understood the general task and the user interface, the experimental session began. During the main part of the experiment, the participants were informed by a screen message about the type of supplementary web pages presented in each session. The browser window displayed a task question, a search term, a supplementary type, and a timer at the top. In addition, under the multiple targets condition the number of already-selected web pages in the current task was provided (Figure 4.2). This information remained on screen during the search task and the subsequent ratings. During each task a beep sounded and a visual sign appeared, if and when, ten seconds remained (Figure 4.1).



**Figure 4.1 Experiment 4: An example of a computer screen presenting forty-eight options (on eight SERPs) with no supplementary web pages**

Figure 4.1 shows an example of an interface of HCI browser during a search task under the single choice selection condition with a 48-option set and no supplementary web pages being provided, with the visual sign indicating there is less than ten seconds remaining as displayed in the top-right corner of the screen.



Figure 4.2 Experiment 4: An example of a computer screen presenting twenty-four options (on four SERPs) with synonym-supplementary web pages

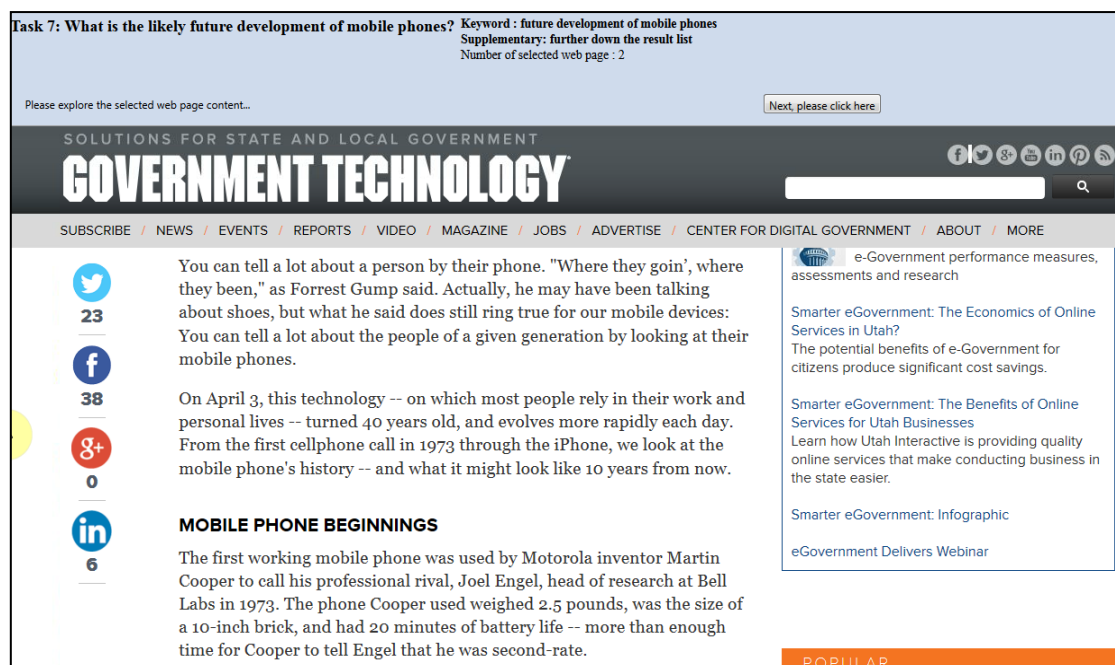


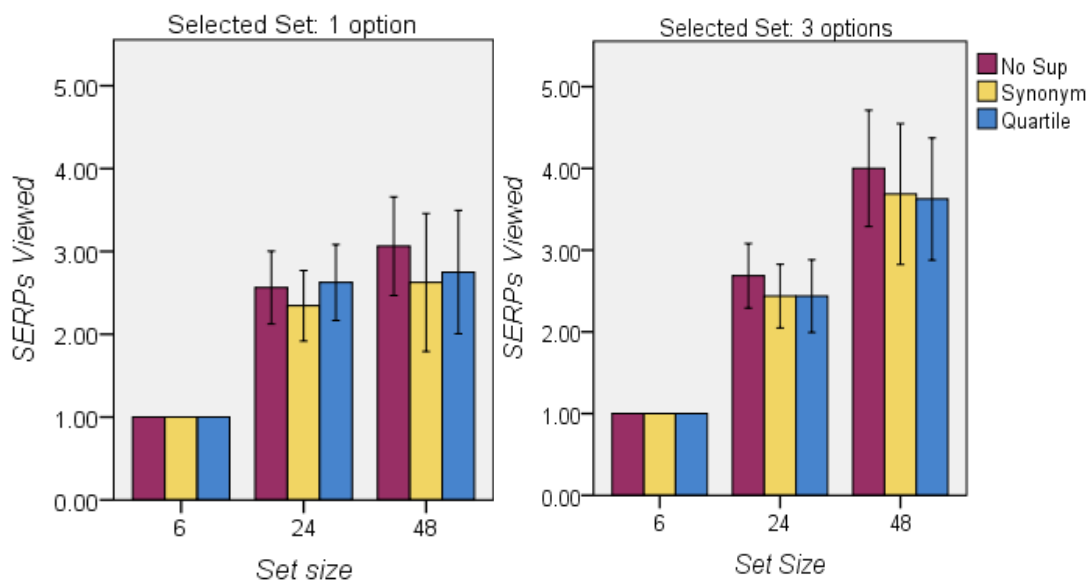
Figure 4.3 Experiment 4: An example of a computer screen displaying a selected web page

Figure 4.2 presents an example of a computer screen during a search task under the multiple choices selection condition with a 24-option set and there being synonym-

search supplementary web pages provided. In this figure, the number of selected web pages is zero, whereby the participant has not chosen any web page for this task. In addition, an example of a computer screen displaying a selected web page is shown in Figure 4.3. In a debriefing session, the purpose of the experiment was explained to the participants, and they were allowed to ask any question related to the study.

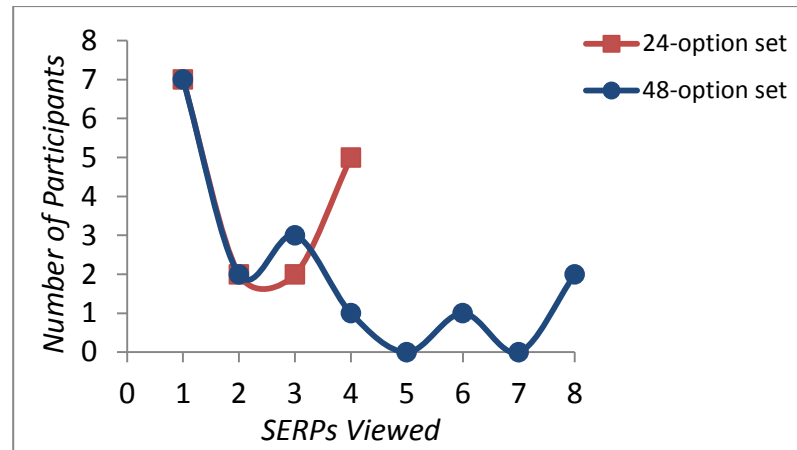
### 4.3.2 RESULTS

#### 4.3.2.1 SERPs VIEWED



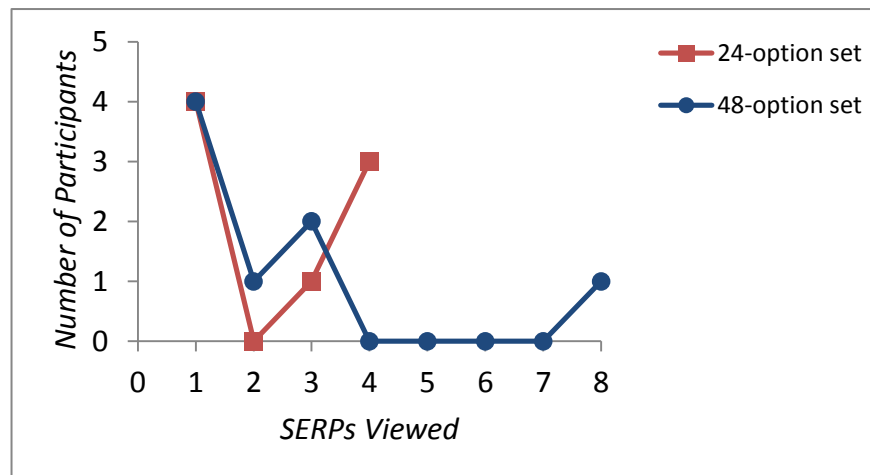
**Figure 4.4 Experiment 4: Mean number of SERPs viewed before selecting for the single selected target (left) and three selected targets conditions (right). Error bars show standard errors.**

Figure 4.4 shows the mean number of SERPs that the participants viewed (out of a maximum of 4 separate SERPs for 24 options and 8 separate SERPs for 48 options) before selecting a web page. According to a one-sample t-test, the number of SERPs viewed in the 24-option set and the 48-option set conditions were significantly greater than 1 ( $t(15)=5.6$ ,  $p<0.001$  and  $t(15)=4.5$ ,  $p<0.001$ , respectively). However, the ANOVAs (2x2x3, Set Size (24 vs. 48) x Selected Set (1 vs. 3) x Supplementary Type) revealed that the main effects of Set Size, Supplementary Type and Selected Set on the number of SERPs viewed in the 24-option set and the 48-option set were non-significant, and there was no significant interaction effect ( $F$ s close to 1).



**Figure 4.5 Experiment 4: Modal number of SERPs viewed per participant.**

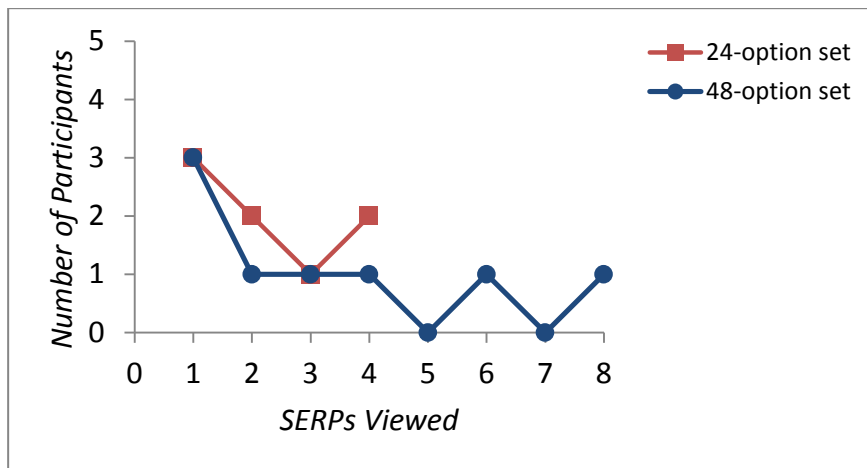
Figure 4.5 shows distribution of the modal number of SERPs viewed by each participant in the two larger set sizes. For both the 24-option set and 48-option set conditions, there was a considerable number of participants who viewed only the first SERP (43.75% in both conditions). In addition, the tendency that the participants would browse all available SERPs in the 24-option set condition was relatively high (31.25%). However, when more SERPs were available, i.e. in the 48-option set, only a few browsed up to the last SERP (12.5%).



**Figure 4.6 Experiment 4: Modal number of SERPs viewed per participant for the single selected target.**

Figures 4.6 and 4.7 separate the data in Figure 4.5 according to the size of the selected set. Figure 4.6 shows the modal number of SERPs viewed per participant for the single selected target condition. In the case where they could select only one option, those who

were presented with 24-option sets tended to browse either only the first SERP (50%) or all available SERPs (37.5%). That is, it was more likely that once the participants had browsed beyond the first SERPs, they tended to look through all the options provided. However, when the number of available options increased, i.e. in the 48-option set condition, participants were less likely to browse all the available options (87.5% viewed up to three SERPs), which suggests that too many options can negatively affect their motivation for browsing more options.



**Figure 4.7 Experiment 4: Modal number of SERPs viewed per participant for the multiple selected targets.**

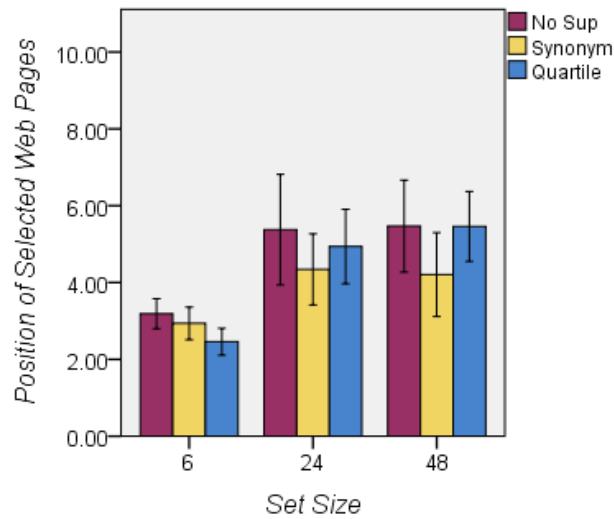
Figure 4.7 shows the modal number of SERPs viewed per participant for the multiple selected targets condition. In this situation, the browsing behaviours were different in that some viewed options only on the first SERP (37.5%), whereas most browsed through more than one SERP.

#### 4.3.2.2 POSITION OF THE SELECTED WEB PAGES

The number of web pages that the participants had to select per task was manipulated between-subjects, either one page or three pages. In this case, the mean ordinal position of selected web pages under that condition was analysed separately. Therefore, the analyses reported in this section are 3x3, Set Size (6 vs. 24 vs. 48) x Supplementary Type (No supplementary vs. Synonym supplementary vs. Quartile supplementary) mixed ANOVAs.

#### 4.3.2.2.1 Single selected target condition

Figure 4.8 shows the mean ordinal position of the selected web pages from the main search result lists (the selected web pages from the supplementary lists are not considered in this analysis). There was a significant main effect of Set Size on the mean ordinal position of the selected web pages ( $F(2,21)=3.85$ ,  $p=0.038$ ,  $\eta^2_p=0.27$ ). The main effect of Supplementary Type and the interaction effect were non-significant ( $F_s<1$ ). Participants in the 24-option and 48-option set conditions selected web pages from further down the list than did participants in the 6-option set condition. Although the majority of the selected web pages were located in the first SERPs, participants mostly checked the results from other SERPs before selecting the web page (Figure 4.4, left).

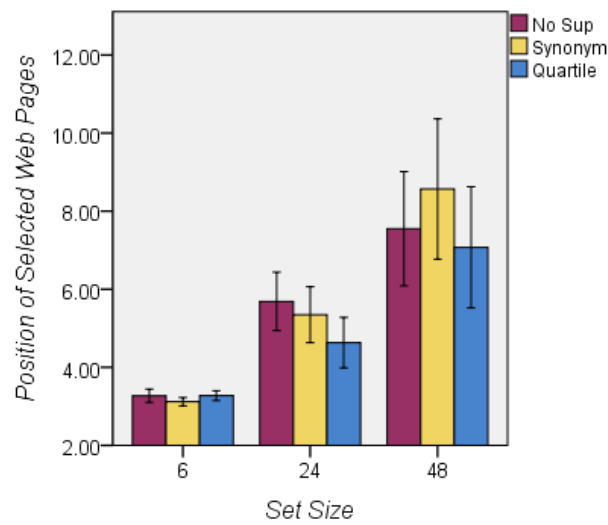


**Figure 4.8 Experiment 4: Mean position of the selected web pages in the single selected target condition. Error bars show standard errors.**

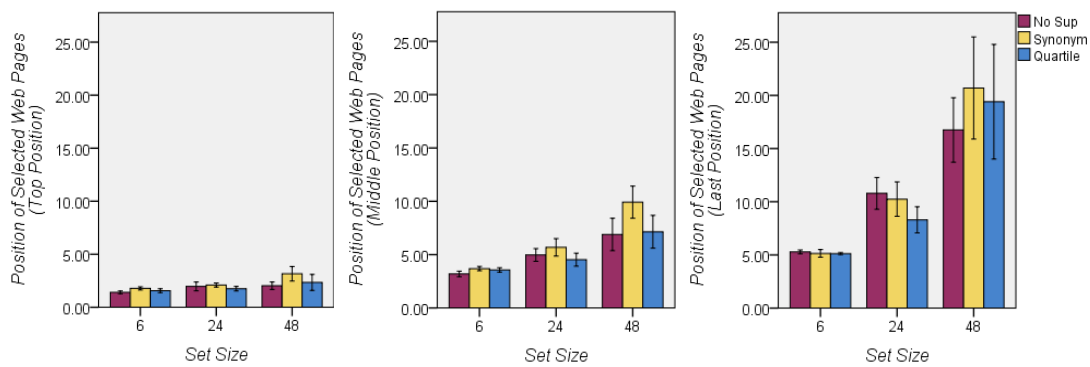
#### 4.3.2.2.2 Three selected targets condition

Figure 4.9 shows the mean ordinal position of the selected web pages in the multiple selected targets conditions (the selected web pages from the supplementary lists are not considered in this analysis). The results indicate that the participants selected web pages from further down the result list when they were choosing from the larger set of options. The main effect of Set Size on the ordinal position of the selected web pages was significant ( $F(2,21)=5.37$ ,  $p=0.013$ ,  $\eta^2_p=0.338$ ). The main effect of Supplementary Type was non-significant ( $F(2,42)=2.97$ ,  $p=0.062$ ,  $\eta^2_p=0.124$ ) and there was no significant interaction effect ( $F$  closes to 1). To focus on the allocation behaviour, the positions of selected web pages from each task were grouped into top, middle, and last.





**Figure 4.9 Experiment 4: Mean position of the selected web pages in the multiple selected targets condition. Error bars show standard errors.**



**Figure 4.10 Experiment 4: Top position of the selected web pages (left), middle position of the selected web pages (middle), and last position of the selected web pages (right). Error bars show standard errors.**

Figure 4.10 shows the mean ordinal position of the selected web pages: top, middle and last. The effect of Set Size on the ordinal top position was non-significant ( $F(2,21)=2.77$ ,  $p=0.08$ ). The top position of selected web pages was mostly on the first SERP (Figure 4.10, left). There was no significant main effect of Supplementary Type on the ordinal top position of the selected web pages and no interaction effect ( $F$ 's close to 1).

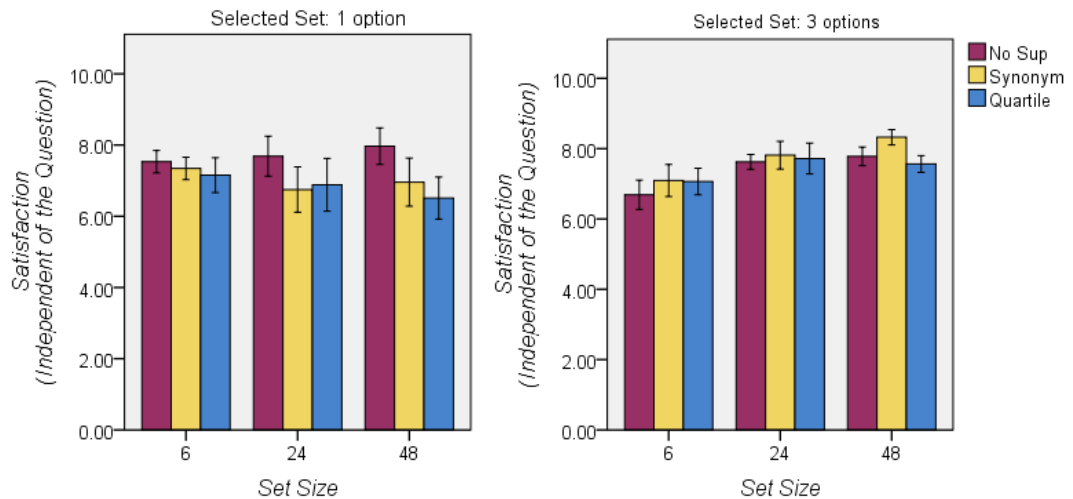
The main effect of Set Size on the ordinal middle position was significant ( $F(2,20)=7.9$ ,  $p=0.003$ ,  $\eta^2_p=0.44$ ) and the participants significantly selected web pages that were located further down the result lists when they were provided with the larger options sets. The main effect of Supplementary Type on the ordinal middle position was also significant ( $F(2,40)=6.7$ ,  $p=0.003$ ,  $\eta^2_p=0.25$ ). The positions of the selected web pages in

the Quartile-supplementary and the No Supplementary conditions were located closer to the top of the results lists than in the Synonym-supplementary condition (Figure 4.10, middle). This suggests that selection behaviour was affected by the perceived quality of the results from further down the list represented by quartile-supplementary web pages. There was no interaction effect between Set Size and Supplementary Type ( $F(4,40)=2.21, p=0.085$ ).

There was a main effect of Set Size on the ordinal last position of the selected web pages ( $F(2,17)=10.42, p=0.001, \eta^2_p=0.55$ ). That is, the participants selected one of the three web pages that was located significantly further down the results list in the 48-option set than in the 6-option set and in the 24-option set conditions (Figure 4.10, right). The main effect of Supplementary Type and the interaction effect were non-significant ( $F_s<1$ ).

#### 4.3.2.3 RATING OF THE SELECTED WEB PAGES

The main analyses reported in this subsection are 3x2x3, Set Size (6 vs. 24 vs. 48) x Selected Set (1 vs. 3) x Supplementary Type (No supplementary vs. Synonym supplementary vs. Quartile supplementary) mixed ANOVAs.

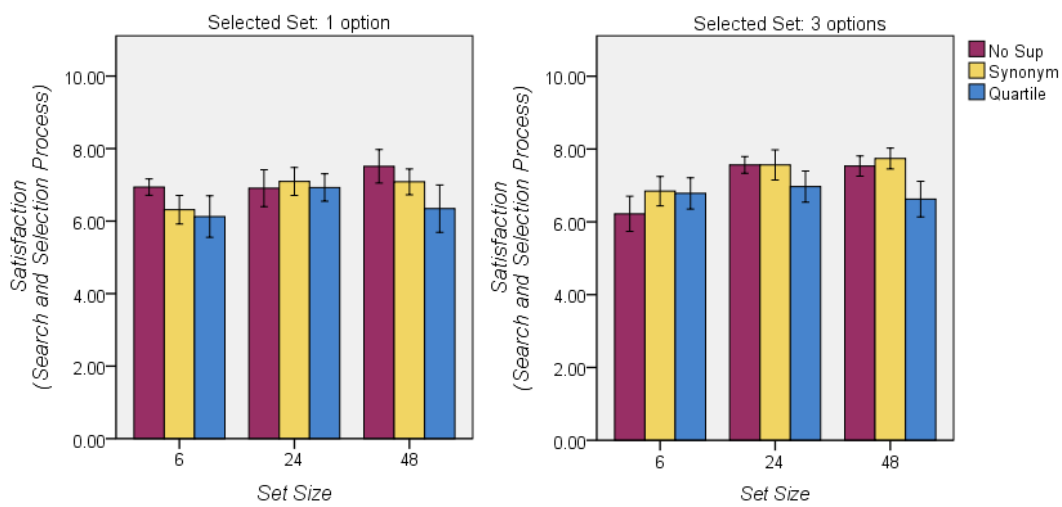


**Figure 4.11 Experiment 4: Mean satisfaction with selected webpages independent of the question, for the single selected target (left) and three selected targets conditions (right).**

**Error bars show standard errors.**

There was no significant main effect of Set Size or Supplementary Type and no interaction effect on participants' mean satisfaction with the selected web pages with respect to the information those web pages provided concerning the task question. (all  $F_s<1$ ). Similarly, there was no significant main effect on the mean satisfaction with

selected web pages when this satisfaction judgement was independent of the task question. However, the interaction effect between Supplementary Type and Selected Set was significant ( $F(2,84)=4.33$ ,  $p=0.016$ ,  $\eta^2_p=0.093$ ). When the participants were required to select a single web page per task, they were more satisfied with their selected options when none of the supplementary web pages were provided. On the other hand, when participants were required to select three web pages per task, they were relatively more satisfied with their selected web pages under the synonym-supplementary condition (Figure 4.11).



**Figure 4.12 Experiment 4: Mean satisfaction with the search and selection process, for the single selected target (left) and three selected targets conditions (right).**

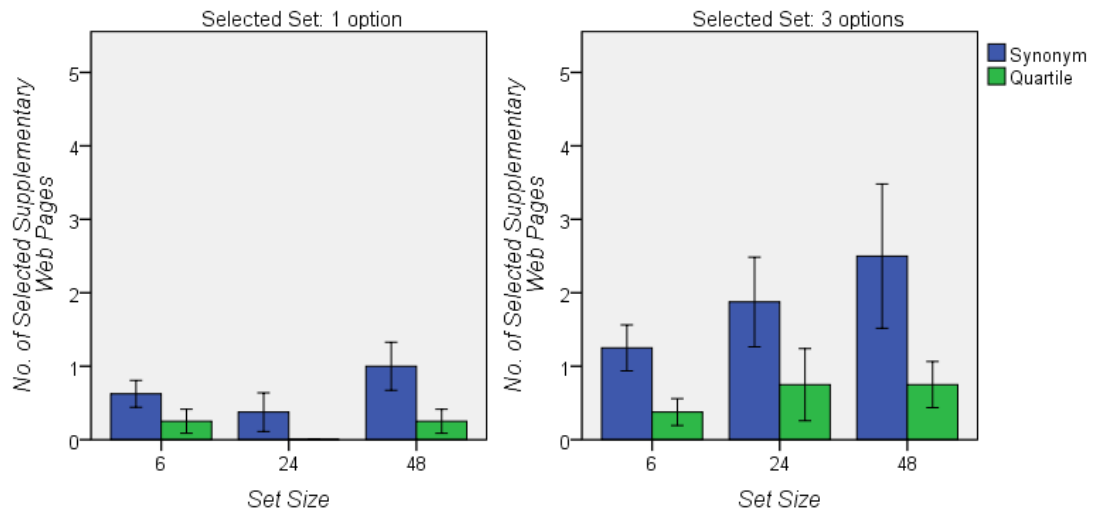
**Error bars show standard errors.**

Figure 4.12 shows participants' rated satisfaction with the search and selection process. The main effect of Supplementary Type was significant ( $F(2,84)=3.86$ ,  $p=0.02$ ,  $\eta^2_p=0.08$ ). That is, the participants were more satisfied with the search and selection process in the Synonym-supplementary and the No supplementary conditions than in the Quartile-supplementary condition. There was no significant main effect of Set Size or interaction effect on participants' satisfaction with the search and selection process ( $F(2,42)=2.54$ ,  $p=0.09$ ,  $\eta^2_p=0.11$ ). There were no main significant effects or interactions on the Satisfaction judgements for the lists of returns (all  $F$ s close to 1). Similarly, no main effects or interactions on Confidence judgement were found (all  $F$ s close to 1).

The Satisfaction with the search and selection process was very strongly correlated with Satisfaction judgements about the selected web pages, in all aspects: satisfaction

judgement concerning the task question ( $r=0.70$ ,  $p<0.001$ ), satisfaction judgement independent of the task question ( $r=0.68$ ,  $p<0.001$ ) and satisfaction judgement for the lists of options returns ( $r=0.75$ ,  $p<0.001$ ). In addition, the Satisfaction with the search and selection process was reasonably strongly correlated with the Confidence judgement ( $r=0.50$ ,  $p<0.001$ ).

#### 4.3.2.4 NUMBER OF THE SELECTED SUPPLEMENTARY WEB PAGES



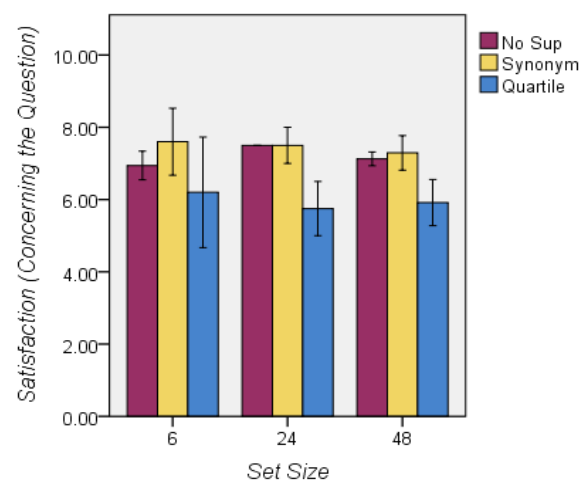
**Figure 4.13 Experiment 4: The number of selected supplementary web pages, for the single selected target (left) and three selected targets conditions (right).**

**Error bars show standard errors.**

The number of selected supplementary web pages was accumulated from twelve search tasks and those from the synonym-supplementary list were selected more than those from the quartile-supplementary list (Figure 4.13). The 3x2x2, Set Size (6 vs. 24 vs. 48) x Selected Set (1 vs. 3) Supplementary Type (Synonym supplementary vs. Quartile supplementary) mixed ANOVAs reported a significant main effect of Supplementary Type ( $F(1,42)=16.7$ ,  $p<0.001$ ,  $\eta^2_p=0.28$ ). The participants chose supplementary web pages from the synonym-supplementary list more than those from the quartile-supplementary list. In addition, the main effect of Selected Set was significant ( $F(1,42)=10.04$ ,  $p=0.003$ ,  $\eta^2_p=0.19$ ). The supplementary web pages were selected more when the participants had to select three web pages per task. However, there was no main effect of Set Size and no interaction effect ( $F$ s close to 1).

#### 4.3.2.5 RATING OF THE SUPPLEMENTARY SELECTED WEB PAGES

The data were pruned such that for the two conditions where supplementary lists were provided, only satisfaction ratings of the supplementary selected pages (for the single selected target condition) or the selected sets that included selected web pages from the supplementary lists (for the three selected choices condition) were used in the analyses. The main analyses reported in this subsection are 3x2x3, Set Size (6 vs. 24 vs.48) x Selected Set (1 vs. 3) x Supplementary Type (No supplementary vs. Synonym supplementary vs. Quartile supplementary) mixed ANOVAs.

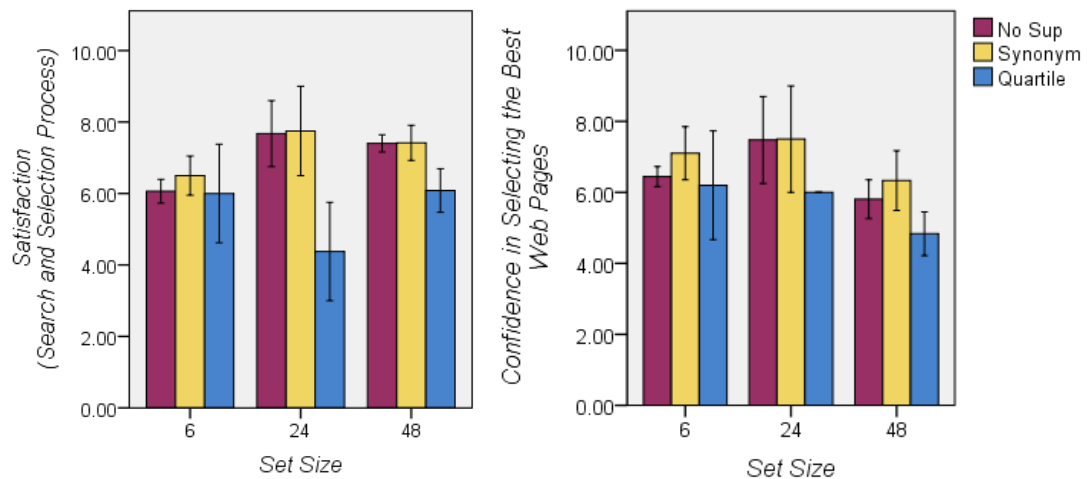


**Figure 4.14 Experiment 4: Mean satisfaction with the supplementary selected web pages concerning the questions. Error bars show standard errors.**

Figure 4.14 shows participants' rated satisfaction with their chosen web pages concerning the task questions. The main effect of Supplementary Type on the mean satisfaction was significant ( $F(2,16)=4.33$ ,  $p=0.032$ ,  $\eta^2_p=0.35$ ). The participants were more satisfied with their selected web pages chosen from the Synonym-supplementary list than those chosen from the No supplementary list or the Quartile-supplementary list. However, there was no main effect of Set Size and Selected Set, and no interaction effect ( $F$ s close to 1). There were no significant main effects or interactions on the Satisfaction judgements for the selected web pages independent of the task questions or on the satisfaction judgements for the lists of returns (all  $F$ s close to 1).

However, the main effect of Supplementary Type on the mean satisfaction with the search and selection process was significant ( $F(2,16)=5.14$ ,  $p=0.019$ ,  $\eta^2_p=0.39$ ) and on the Confidence judgement ( $F(2,16)=3.7$ ,  $p=.048$ ,  $\eta^2_p=.32$ ). The participants were more

satisfied with their search and selection process as well as more confident in their selected options in the Synonym-supplementary and the No supplementary conditions than in the Quartile-supplementary condition (Figure 4.15). In addition, there was no significant main effect of Set Size and Selected Set, and no interaction effect on these two ratings ( $F$ s close to 1).



**Figure 4.15 Experiment 4: Mean satisfaction with the search and selection process (left), and confidence in selecting the best web pages (right). Error bars show standard errors.**

#### 4.3.2.6 OVERALL RATING OF THE SELECTED WEB PAGES

The participants were asked to rate their overall satisfaction at the end of each task-session (i.e. after each supplementary type). There were no main effects or interactions on overall satisfaction for selected web pages and for the lists of options returned, with all  $F$ s being close to 1.

#### 4.3.2.7 STRATEGIES OF CHOOSING A WEB PAGE

The participants responded to the open-ended question at the end of each task-block (i.e. after each supplementary type) about their strategy for choosing a web page and the usefulness of the supplementary web pages provided. This qualitative analysis was performed in order to find insights about search behaviours and preferences. All responses were transcribed by the researcher, with each transcripts being read twice and then analysed using thematic analysis (Braun & Clarke, 2006). The transcripts were given to an independent coder in order to be analysed for the presence or absence of the identified themes so as to ensure the reliability of the coding. Most of the participants reported several search strategies used for each task-block, which were classified into

five main themes: *validity of known sources*; *indication of information details*; *keyword matching*; *order of options*; and *unfamiliar or unknown sources*. These themes presented the basic strategies that participants used, separately or in combination, during the experiment.

First, *validity of known sources* was mostly reported as a primary concern during the search process. That is, the participants mostly opted for the recognised web sites that were considered, by their content features to be reliable and trustworthy sources of information. Some mentioned that they use the known characteristics of the web site content to verify its validity and when this was the case it would be deemed suitable for answering a particular question. In addition, two sub-themes identified within *validity of known sources* were *reliability of sources* and *suitability of the context*.

The first sub-theme, *reliability of sources* was mainly considered by most of the participants during the search processes. Reliability judgement was based on participants' individual preferences and recognition from past experiences. That is, recognised web sites were used by some participants to judge the reliability of the particular pages.

[Participant #3, Male, Undergraduate, 6-option set, Single choice] *"I used web sites that I have used in the past, to ensure the links were reliable enough."*

[Participant #35, Male, Undergraduate, 24-option set, Multiple choices] *"Choosing the websites I recognised and have possibly used before"*

[Participant #39, Male, Postgraduate, 48-option set, Multiple choices] *"Recognising what websites are likely to provide useful information is important. Some websites seem to return keywords which match the question's keywords but actually do not provide too much useful information"*

The second sub-theme, *suitability of the context* referred to statements where the participants focused on the context of the task question and tried to find a suitable answer from known web sites. Examples of such responses are as follows:

[Participant #2, Female, Undergraduate, 6-option set, Single choice] *"Apart from the Scotland question, where I chose trip advisor because I wanted a word of*

---

*mouth feel, I chose web pages that seemed to originate from credible and knowledgeable sources.”*

[Participant #25, Male, Postgraduate, 6-option set, Multiple choices] *“Examined the titles of the pages and if needed I read the text under the title. I also looked at the URL to see if it was from a recognised site that I thought might answer or not answer the question – for example Wikipedia is useful for a general overview and pointers, but Wikianswers is usually someone asking a question and not always getting a response”*

[Participant #45, Male, Undergraduate, 6-option set, Multiple choices] *“I would avoid Wikipedia for ‘intellectual’ searches. I prefer forums of mediums like ask.com or yahoo ask where people post questions and other people answer the question...”*

[Participant #38, Female, Postgraduate, 48-option set, Multiple choices] *“1.title 2.source 3.year of publication (just for the future of mobile question) I generally trust .gov sites and I use blogs for tech matters (comments can be useful!!)”*

Second, *indication of information details* pertained to statements where a website was selected based on the information it might provide. Some participants preferred a different level of information details compared to the others. That is, some preferred a web page that summarised information, for example, looking for the numerical points in the title of search results, whilst others preferred a web page that provided information in more detail. For example:

[Participant #36, Male, Postgraduate, 24-option set, Multiple Choices] *“Choosing web pages that give information rather than those answering a single question.”*

[Participant #44, Male, Undergraduate, 6-option set, Multiple Choices] *“Look for some words like ‘eight strategies’, ‘three steps’. Do not only look for the explanation of the general concept”*

Third, *keyword matching* referred to the strategy that the keywords from the task question were used to identify the relevant titles or snippets of search results. Some



participants relied on keyword matching more than the others, and less focused on the reliability and suitability of sources. For example:

[Participant #6, Female, Postgraduate, 6-option set, Single Choice] *“I tried to find the keywords I have written on the link then very briefly read each explanation in each link.”*

[Participant #18, Female, Postgraduate, 48-option set, Single Choice] *“I tried to read a snippet below the actual web link, and I also tried to compare the wording of the heading with the question to obtain the most accurate result.”*

[Participant #27, Female, Postgraduate, 6-option set, Multiple Choices] *“I read the headings first, checking which was closest to the question. Then I read the info given under the heading. If it still fitted the question, I clicked on the link.”*

Next, *order of options* was used mostly in combination with other strategies. Some participants relied on the ranking algorithm, such that they believed that the most relevant results were presented on the first SERP. For example:

[Participant #8, Female, Undergraduate, 24-option set, Single Choice] *“My strategy involved reading the title and descriptions of the website, particularly in the first pages as they seemed most relevant. I tried to pick out websites that avoided answers by the public, such as Wikianswers and Wikipedia.”*

[Participant #45, Male, Undergraduate, 24-option set, Multiple Choices] *“...I tend to choose options from the first page, and have least interest in the last page. Usually, I would change my search if the web pages shown in the first few options are not helpful”*

[Participant #48, Male, Undergraduate, 48-option set, Multiple Choices] *“Most of the time, I only pick the links from the first result page and always pick the second link to begin with. By looking at the website URL also helps me to decide to choose the links.”*

In addition, the effect of ranking was reported when one participant chose the web page from further down the results list. She seemed to regret doing so, even though at the time of selecting she thought the web page was the most suitable option for the task question:

---

[Participant #21, Female, Postgraduate, 24-option set, Single Choice] *“For the last question I selected the link from the last 4<sup>th</sup> page, as I couldn’t find anything suitable earlier, but because of it I do not feel sure if it was the best possible choice.”*

Finally, one participant mentioned that one of the three selected web pages for some tasks would be from *unfamiliar or unknown sources*, which shows behaviour aimed at trying to increase the variety of the choice set. The response he gave was as follows:

[Participant #40, Male, Postgraduate, 48-option set, Multiple Choices] *“... I went to the most reputable and well established website, although sometimes I tend to select at least one website which I am not familiar with.”*

#### **4.3.2.8 USEFULNESS OF THE SUPPLEMENTARY WEB PAGES**

The participants were asked to respond to the open-ended question at the end of each task-block about the usefulness of each type of the supplementary web pages. Based on the dataset, the participants seemed to favour supplementary web pages from the Synonym search more than from the Quartile search. This finding corresponds, in general terms, with the quantitative judgements reported above. In addition, the usefulness of supplementary web pages from the Quartile search was reported as being unpredictable, because often the snippets did not represent the actual webpage content. For example:

[Participant #13, Female, Postgraduate, Quartile-supplementary] *“The performance seems quite dynamic, sometimes really helpful, sometimes are not. I slightly prefer the supplementary from last session (Synonym supplementary) than this.”*

[Participant #6, Female, Postgraduate, Quartile-supplementary] *“It was quite useful to predict the content I would have later, but sometimes it was a bit disappointing not to meet with my expectation”*

The perceptions regarding these pages’ usefulness differed. In the Synonym-supplementary condition, some participants reported that these pages somehow changed the interpretation of the task, whilst other said that they provided general background

about the task question. The responses were classified into three main themes: *additional information*; *interesting search results*; and *misleading*.

First, as *additional information*, supplementary web pages from both Synonym and Quartile searches provided some additional information that was directly or indirectly related to the search task. Examples of responses regarding this were as follows:

[Participant #17, Male, Postgraduate, Synonym-supplementary] “*For the question which I have background knowledge on, the supplementary web pages were not useful. However, if I didn’t have any knowledge on that, it was useful.*”

[Participant #25, Male, Postgraduate, Synonym-supplementary] “*It was quite useful when it did provide alternative search results...*”

[Participant #28, Female, Postgraduate, Quartile-supplementary] “*They did not seem to be directly addressing the questions, more as additional information*”

[Participant #39, Male, Postgraduate, Quartile-supplementary] “*Most of the supplementaries seem to provide some answer to the question, but not exactly the good answer I was looking for. It’s useful if I want to know more information providing that I have already found other better answer, i.e. they can be used as additions.*”

[Participant #39, Male, Postgraduate, Synonym-supplementary] “*The quality of the supplementries is better than the first session. A synonym is useful in searching.*”

Second, *interesting search results* pertained to statements where the supplementary web pages were somewhat interesting, for example:

[Participant #16, Female, Postgraduate, Synonym-supplementary] “*They mostly provide a more global context surrounding our question, in that sense they are interesting. For really specific questions I didn’t select them.*”

[Participant #42, Male, Postgraduate, Synonym-supplementary] “*Quite good. They provided websites that have some interesting titles.*”

Finally, *misleading* referred to statements where supplementary web pages were seen as irrelevant to the tasks, for example:

---

[Participant #28, Male, Postgraduate, Synonym-supplementary] *“It’s sometimes irrelevant to my question but can provide useful and more accurate results generally.”*

[Participant #40, Male, Postgraduate, Synonym-supplementary] *“In the first questions the supplementary was useful since it provides a complement to the field of the question. In the other two cases, I think the supplementary was misleading from the question’s main purpose.”*

#### **4.3.3 DISCUSSION**

The results from this experiment revealed no evidence of a choice overload effect in the satisfaction ratings of the selected web pages, and indeed the non-significant tendency was that the participants were more satisfied with the large set of search returns, which is broadly consistent with the findings from Chapter 3 and shows that the lack of a choice overload effect is not a matter of experimental power. This is true even though the time constraint of 45 seconds is quite severe for browsing the 48-option set, this did not lead to significantly less satisfaction.

Time constraint might be the reason that the number of SERPs viewed in the 48-option set condition was less than half of the total number of SERPs available. Participants in the 24-option set and 48-option conditions browsed through options from further down the SERPs (average SERPs viewed is about 2.5 and 3, respectively). However, in the multiple selected targets condition where the participants’ selection time was triple, most of them browsed through a few SERPs in the 48-option condition. Moreover, most of the selected web pages in those conditions were located on the first SERP. Again, the participants’ behaviour shows that their choice of selected web pages was influenced by the ranking of returns, for they mostly relied on the ordering of options and seemed to be satisfied with the selected options even when they could not verify most of those available.

It is worth noting that, the number of selected supplementary web pages was relatively few compared to the number of web pages selected from the Google results list in the single selected target condition (Figure 4.13, left). In the 48-options condition, about one supplementary page over all 12 tasks, which could have been because these were viewed as distractions. That is, the participants might have perceived that the supplementary web

pages were diverting them away from the tasks and thus, decided, on the whole, to ignore them (Tom et al., 2009).

### **4.3.3.1 SINGLE TARGET SELECTION**

When the satisfaction was judged independent of the task question, the participants who had to choose only one option per task were less satisfied with their selected web pages when supplementary lists were presented, even though the majority of the selected web pages were chosen from the Google results list across all supplementary type conditions. In this case, regret about rejecting the opportunity to explore options from the different results list perhaps affected the satisfaction with the selected options. This might possibly be due to the paradox of choice in that providing the participants with extra options meant that they had to forego selecting some that may have brought reward. As a consequence, they were less satisfied than if they had not been offered such variety in the first place. That is, not knowing about something means not missing it. In addition, when no supplementary option was available they could be comfortable in trusting the ranked list without having to think about other things that might negatively impact on their satisfaction level. However, it is also important to note that the above outcome regarding supplementary lists may be due to the limitation that only one option was to be selected.

### **4.3.3.2 MULTIPLE TARGETS SELECTION**

In this experiment, search behaviours in the context that required multiple targets were additionally investigated. This was to explore users' search strategies and satisfaction when variety of options could be one of the search criteria. The tasks used in this experiment were open-ended problems, which allowed the participants to either select each option from a different aspect or choose options that were similar in content if they desired. The findings revealed that when multiple options were required to be selected for each task, the participants' variety seeking behaviour confirmed that the single ordering of search results indeed affects the perceived variety among those returns, which leads to adaptation of the search and selection process. For instance, participants acted against the implication that items from further down the ordered list would be less relevant to the task. However, the results showed that in each task all the participants took at least one of their three attempts to choose a web page from near the top of the result lists (mostly on the first SERP). Nearly all, followed this looking for and selecting

---

web pages from further down the list or opting for supplementary web pages. Even though further down the ranked lists it was quite clear that the relevance of a web page to the topic at hand was increasingly diminishing, some still chose right towards the bottom, even in the 48-option set condition. Thus, it seems that the effect of ranking of returns is less significant in multiple-choice decisions due to the fact that the user tends to increase variety among the selected options. These findings from the experiment are consistent with previous work (Fox et al., 2005; Benartzi & Thaler, 2001), which demonstrated that consumers tend to have bias towards even allocation when choosing products or services based on their subjective partitioning. In addition, the results have extended those works by reporting examples of such behaviour even in the situation where the options are ordered, thus demonstrating that even when participants are informed that certain options are likely to be inferior they still prepared to choose them.

It is interesting that participants in the larger set condition selected web pages from the supplementary lists more than those with the smaller set. Choosing supplementary web pages implies the intention to increase differences among the selected options. It could be assumed that the options in a smaller ranked set are similar in content. However, most of the participants in the small set condition did not choose to increase variety by selecting a supplementary web page. In fact, the number of chosen supplementary web pages in the 48-option set condition was more than those in the 6-option set and 24-option set conditions, but this effect did not reach significance. This may hint at a choice overload effect that led to the participants choosing the supplementary web pages instead of going through the long results list. In addition, the participants who had difficulty in choosing options from a large set may have applied a simple heuristic strategy by selecting the supplementary web pages in order to increase variety (Simonson, 1990).

The participants reported greater satisfaction with the search and selection process in the Synonym-supplementary condition than they did in the other two conditions when they had to select multiple options. The reason may be that this list increased the perception of variety among the available option sets. According to the qualitative analyses, the participants reported that the opportunity to choose three web pages allowed them to select those that were considered to be somewhat different in content. This shows that the perception of variety among the options can increase satisfaction with the decision process. Furthermore, those who tended to seek variety were more satisfied with their selected options from the Synonym-supplementary lists, concerning the task questions.

As results reported, when the sets of selected options that contained at least one page from the supplementary lists were considered, the participants were relatively more satisfied with their selected web pages from the Synonym-supplementary lists than from the other two lists. This suggests that the opportunity to have a variety of options from the Synonym-supplementary list can increase the satisfaction judgement on the set of selected options, given that the judgement was made based on the task questions.

### **4.3.3.3 QUALITATIVE FINDINGS**

In addition to the keyword matching strategy, the very basic one used in the search and selection process, the participants seemed to rely on the order of options and their simple recognition. Thus, it would appear that simple recognition could affect the users' judgements about the reliability of information sources and the suitability of the information context.

Participants' reliability judgement of information sources was in part based on the known quality of a web site's domain name and in part on the recognised reputation of the web site provider. In addition, the level of reliability was judged differently based on individual preference, the prior experience of using a particular web site, and the judgement of content suitability for a specific search objective. For example, when the task was to find a place to travel, some participants treated a web site that provided information from official tourism organisation as highly reliable, whereas others perceived one that provided tourist attraction reviews from unofficial sources as more reliable. This suggests that some search engine users may choose the selected web pages by applying their simple recognition in one way or another in order to make a decision, even in an unfamiliar information context. For example, they may select a web page from a recognised web site, a recognised web site provider or organisation, or a recognised domain name.

The strategies for selecting a web page used in the multiple targets decision context were mostly similar to those used in the single target decision, with the main exception being that some different strategies were employed so as to facilitate variety seeking. For instance, some participants mentioned their preference of selecting a web page that organised its content in the numerical points style, which may facilitate the variety seeking in that it helps users identify the differences between web pages. Moreover, the

---

use of unfamiliar or unknown sources was mentioned, which suggests the intention to increase variety among the selected-option set.

On the usefulness of supplementary web pages, the participants seemed to be more satisfied with those from the synonym search than from the quartile search. The experimental design might have affected their judgement about the quality of supplementary types in that the participants were told about the sources of supplementary web pages during the introduction. This could have affected the relevance judgement of supplementary web pages from the quartile search, especially if the participants relied on the ordering of returns. That is, the web pages located in the third and fourth quartiles of the results list may have been perceived as less relevant based on their location on the ordered list.

The responses about the usefulness of the supplementary web pages were analysed and the results revealed both advantages and drawbacks regarding the supplementary lists. Some participants stated that these web pages from both Synonym search and Quartile search provided some addition information that was directly or indirectly related to the task, i.e. some results were useful or interesting. However in some cases the supplementary web pages were seen as irrelevant to the task and even considered as misleading by some.

#### **4.3.4 LIMITATIONS AND FUTURE WORK**

In this experiment, the participants were asked to find a web page that provided valuable information for each task question. Although the questions used were based on the open-ended problems, they could be classified as goal-directed tasks and this could have been one of the reasons that in most cases the supplementary web pages were not selected. As suggested in a previous study (Toms & McCay-Peet, 2009), those pages may have been viewed as a distraction to achieving the goal. The design of a study that aims to explore chance encountering may thus need to be based on a still more open search task, that is, a search task that naturally involves exploring and discovering may allow a researcher to investigate and gain a better understanding on users' behaviour and their serendipitous experience. The study presenting in Chapter 5 is designed to overcome this limitation.

One limitation of this experiment is that the participants were not asked about their intention to increase variety among the selected options when they had to select three



options per task, i.e. whether they intended to select web pages that provided similar content or were going to choose those that provided different content relative to the task question. Furthermore, they were asked to rate the satisfaction of selected web pages as a set, which meant that what they reported was an average satisfaction with their three choices and hence, no information was available on an individual basis. Consequently, it was not possible to ascertain whether selected pages from further down the list or those found through a synonym search exhibited greater or lesser contentment than those at the top of the list.

The findings suggest that the perception of variety among the set of returns from a synonym search can increase the users' satisfaction with a search and selection process when multiple options are to be selected and hence, it is proposed that search engine design can be improved in order to support the variety requirement by providing search results from different aspects related to the topic of interest. Regarding this, research on search result diversification has mostly tried to capture users' intent by generating a list of results or recommendations that covers many aspects of their information needs (Drosou & Pitoura, 2010; Yu et al., 2009). However, the evidence from this experiment suggests that the presentation of such diversified or what could even be considered random results often leave the searcher confused as to what is relevant. Consequently, when introducing variety, it is recommended that within search terms there is a mechanism that provides a set of categories, say at the top of the screen, informing the searcher of the different ways the subject matter can be interpreted.

### 4.3 CHAPTER SUMMARY

In this chapter, the method and outcomes of an experiment studying search engine users' behaviour have been presented. The possible effect of choice overload on the satisfaction with the selected options was investigated. Moreover, the extent to which variety of search returns affects the variety seeking behaviour and chance of discovering serendipity was examined. Both quantitative and qualitative data were collected and analysed, which have provided a better understanding on user search behaviour and satisfaction. It was found that:

- There was no reliable evidence indicating a choice overload effect in the satisfaction judgements of selected web pages, indeed a large number of

---

options tended to be associated with more satisfaction with the selected options.

- In the situation where a single option was to be selected, reliance on the order of the search results list may have reduced the chance of encountering serendipity in search engine use in that the related search results beyond the ordered list returned were less likely to be considered.
- The perception of variety among the options can increase satisfaction with the search and selection process, in that the users' satisfaction with their selected options may be affected by variety among options rather than a number of available options.
- When variety is required in an ordered environment, the single ordering of search returns influences search engine users to adapt their search strategies and rely less on the results' ordering.
- Users' satisfaction judgements on the search and selection process were strongly correlated with that regarding the selected options, which implies that the satisfaction with the search process could influence that for search outcomes.

In the next chapter, the aim is to explore further chance encountering and serendipity experience in a literature search context. In addition, the extent to which such a search can be supported by an automated synonym-enhanced search is investigated.



## CHAPTER 5

# SERENDIPITY AND LITERATURE SEARCH

### 5.1 CHAPTER OVERVIEW

In Chapter 4 it was suggested that, in the single target decision situation where a search task is considered to be goal-directed, a chance of encountering a surprise discovery or so called serendipitous experience may be limited by searches that focus directly on finding the targeted results. Furthermore, the findings illustrated that searchers are somewhat more likely (although not *very* likely) to select supplementary web pages from the synonym search compared to the quartile search and that these may sometimes provide interesting results; directly or indirectly related to the search topic. The work in this chapter builds upon the potential benefits from the synonym search, with the aim being to explore chance encountering when the ultimate goal of the search is less definite than in the open ended but nonetheless specified tasks so far studied. More specifically, the potential benefit of search results generated by a synonym search term is examined by using a qualitative approach, with the purpose being to address the following research question – **RQ4: How can an automated synonym-enhanced search support a literature search process and increase the chance of encountering a serendipitous experience?** The users' preferences and requirements in literature searching are explored in detail in order to provide a deeper understanding of the potential benefits of the synonym search.

For this study, twelve research students were asked to search for publications that related to their own research area by using the automated synonym-enhanced search application. The task was to find one publication per day for five consecutive days. On the last day of

the experiment, the participants were interviewed about their general use of search engines for research purposes and their experience of searching using the automated synonym-enhanced search application. The data reveal the information needs and moderators that affected literature search behaviour. Regarding serendipitous experiences, the findings suggest that there may be some potential for the synonym search to increase the chance of encountering unexpected information through the retrieval of different aspects of search results.

---

## 5.2 INTRODUCTION

Information seeking is an essential mode of gathering relevant work, evidence, and support for academic research, which is widely supported by search engines in many ways. Scholars' information-seeking behaviour was probed by Ellis (1993) who interviewed academic researchers in social sciences, sciences and humanities in order to model information seeking patterns. This researcher found that fundamental seeking behaviour included the following six characteristics: *Starting*, referring to activities involving the initial search, such as identifying a key publication or using an online search system to locate the main references; *Chaining*, involving the following of chains of citations or other forms of referential connection between publications or other materials; *Browsing*, which refers to semi-directed searching in the area of potential interest in order to identify relevant materials; *Extracting*, which pertains to the act of systematically working through a relevant source, e.g. the same publishers' lists or journals, to find material of interest; *Monitoring*, which denotes the act of regularly following particular sources to maintain awareness of developments in the field or the area of interest; and *Differentiating*, which concerns filtering the quality of materials by examining the differences in the nature of the source materials, verifying the correctness of the information, and prioritising sources based on perceptions of their relative importance.

Search engines play an important role in facilitating these information-seeking activities. Deeper understanding of the requirements underlying the use of search engines in this context is vital for their improvement, which in turn will enhance academic information seeking. In addition to the activities listed above, another that is valuable in the context of academic information seeking is accidental discovery or so called "serendipity". The meaning of serendipity taken from the Oxford English Dictionary (Dictionary, 2013) is "the occurrence and development of events by chance in a happy or beneficial way". It was found that a number of important discoveries in science have come about by accident (Roberts, 1989; Rosenman, 1988), which suggests that increasing the likelihood of serendipity may yield more benefits to research than were it otherwise. However, the occurrence of serendipity is somewhat unpredictable and its nature in the information-seeking context is still open to debate, which consequently, means that further investigation of the phenomenon is required.

With regards to browsing, the basic action in the information seeking process, it would appear to be an important source of creative ideas and chance encountering. Researchers have identified three common forms (Apted, 1971; Herner 1970): (1) specific browsing involving directly browsing for a specific literature, (2) general purposive browsing pertaining to planned or unplanned examination of academic sources and (3) general browsing, where undirected, random, non-purposeful or passing-time browsing is engaged with. Rice et al. (2001) suggested that these different forms of browsing have emerged according to four dimensions, which were later identified as the basic features of encountering serendipitous experience (Foster & Ford, 2003). Firstly, there is *the act of scanning*, where searchers use their existing knowledge of a topic and their understanding of data formats to direct attention when looking, examining, or sampling. Secondly, there is *the presence or absence of purpose*, which directly influences the act of browsing. The third aspect is *the specificity of search outcomes or goals*, which refers to the level of specificity of the search criteria regarding the object being sought and this directs browsing intensity. Finally, *knowledge about the resources and object sought*, involves browsing being directed by the searchers' experience of the resources including their prior knowledge of search paths and/or the content of resources.

In an exploratory study by Nutefall and Ryder (2010), faculty staff and librarians who had collaborated in the past academic year were interviewed about their own research projects. During the interviews, the participants described the instances of encountering unexpected materials that guided their projects in new directions. Three kinds of serendipitous research occurrences were identified: chance, mystery investigation and browsing. Chance was described as the moment when interesting connections are come across while examining unusual sources as potential contributions to the research project. Whilst, mystery investigation refers to learning from detecting, whereby bibliographies or citations are drilled down for additional sources and the final kind, browsing, involves a casual search for items of interest without clearly defined intentions.

Despite the above findings that general browsing without a specific purpose can increase the chance of encountering serendipity, researchers have suggested that individual differences may affect the encountering of such experience (Foster & Ford, 2003). In particular, existing knowledge or creativity of an individual can lead to them drawing connections between unrelated pieces of information that others may not see. In a study

---

of serendipity and information seeking, Foster and Ford (2003) explored different aspects of the information encounter and the generation of new ideas among interdisciplinary researchers. The results demonstrated that certain attitudes and strategic decisions were perceived to be essential for a serendipitous experience to happen. Moreover, the participants in this research perceived that some element of control should be exercised to increase chance encounter, for example, the information gatekeepers could facilitate a serendipitous experience by recommending what the researcher should know or notifying researchers about the possible connection between their existing body of research and other relevant academic work.

Returning to the concern about chance encounters in online environments, keyword searching that directs searchers in a targeted manner, understandably, put limits on a chance encounter. Martin and Quan-Haase (2014) conducted a study interviewing twenty academic historians about their use of e-books and the role of serendipity in their research. The findings revealed that the physical library was the preferred place for searching and browsing for material. The primary concern about chance encountering in the online environment was that search engines and digital documents would facilitate the search process by providing quick retrieval of targeted information. During the aforementioned interviews, the researchers generally reported the view that when they used such techniques they were not expecting them to provide information beyond their specific target and consequently were unlikely to come across anything by chance. However, it was acknowledged that with the efficiency of online searching was it almost inevitable that relevant things would emerge that could not be elicited from a conventional physical search. A number of the participants stated that they were willing to experiment with e-books and anticipated that serendipity could be introduced via an improved system, which provided extra functions to support browsing than at present.

While the studies reviewed above have identified the potential instances that can lead to the encountering of serendipity in information seeking, no work has yet examined how search engine designs could be improved to facilitate this in the context of literature searches. Instead, most have tried to identify the research activities engaged in at the time of encountering serendipity and the types of unexpected information retrieved. Thus, it would be necessary for research to investigate the potential tools or designs that could support serendipity encountering and could be prove beneficial, especially with regards to the real-world literature search context.



Research has demonstrated that qualitative approaches can elicit genuine perceptions of individuals regarding specific situations or environments (Reneker, 1993; Rieh, 2004; Vandenbosch & Huff, 1997). In particular, this form of research is appropriate for studying the needs underlying information seeking activities (Ellis, 1993). In the work presented in this chapter, we are aiming to provide insights into academic search behaviour and examining the use of synonyms in literature searches. Our study focuses on understanding the participants' perception of their search experience rather than deriving a model of seeking behaviour. The findings were achieved through semi-structure interviews that were conducted after the participants had performed search tasks using the automated synonym-enhanced search application. Participants' intrinsic motivation was encouraged by allowing them to perform search tasks based on their own research topic. We hypothesized that evidence of the benefit of synonym search with regards to enhancing the opportunities to experience serendipity would be found, thus increasing the users' satisfaction.

In summary, then, the study reported in this chapter attempts to evaluate a simple design idea that might increase the chances of serendipitous discoveries during search engine use for literature research. The very idea of designing for serendipity is paradoxical to some extent, as serendipity implies accident, which might seem impossible to design for. The idea of an automatically generated synonym-based parallel search attempts to square this circle, by generate search returns that are connected with the users' behaviour, but to some extent out of the user's control – rather as the books on adjacent shelves in a library might be. The way synonym search has been implemented is rather simple, with little grammatical processing or sophistication; but it is unclear whether or not this is a weakness in the design as it perhaps adds to the uncertainty of the synonym-generated search. Nevertheless we hope that a qualitative evaluation of the idea will allow some assessment of the specific design as well as the general principle. It is anticipated that searching by using a synonym may yield useful search results that are unexpected by a searcher, especially when a synonym is unknown.

Another problem for design *evaluation* is raised by the nature of serendipity — serendipitous events are inevitably rare, and one can hardly expect to discover many of them in a manageably small-scale user study. This limitation is recognized, but seems unavoidable; the qualitative nature of the study in this chapter at least allows rare events to be noted and discussed.

---

The study presented in this chapter attempted to explore this by implementing a simple application that would illustrate an example of design towards serendipity. In addition, a number of extra search results generated from a synonym search may also affect the search and selection behaviour, such that more options presented in a list may engage users to browse through more extra results and think about possible connections between these options and their topic of interest.

## **5.3 STUDY 5: AUTOMATED SYNONYM-ENHANCED SEARCH FOR SERENDIPITOUS EXPERIENCE**

### **5.3.1 METHOD**

#### **5.3.1.1 DESIGN**

The study was divided into two parts. The first was designed so that the participants carried out a literature search using the provided application for a period of five consecutive days. The later part involved semi-structured interviews about participants' experiences of using search engines for literature searches in their daily life and about using the provided application to perform the literature searches for the study. In more detail, for the first part of the study each participant completed five literature-search sessions (one session per day). For each session, they were required to select an article related to their own current research topic that they had never read before. They were informed that the expected time spent for each session was about thirty minutes, however, they could spend longer if they so wished.

The participants performed real-time Google searching with a modified interface for scholar articles by specifying their own search terms. In addition, they were informed that every search term they entered would be extended with "PDF" to favour articles as search returns. Moreover, they were allowed to iterate their search terms and open any articles as many as they wanted from SERPs with no time limitation. In order to complete a session, the participants had to submit one article that was considered to be the most valuable for their search goal. During the search task, participants were provided with two lists of search results, which were displayed side by side:

- *Keyword-search list*: Participants were informed, accurately, that the results list had been generated from their specified search term by using Google web search.

- *Synonym-search list*: Participants were informed, accurately, that the results list had been generated by a synonym search term, which was formulated by substituting the first noun found in the participant's search term with a synonym before searching using Google web search.

There were forty-eight options constantly displayed in the keyword-search list (on eight SERPs, reached by pressing a page number button at the bottom of the list). However, the number of options in the synonym-search results lists was manipulated between-subjects: two options (on a single SERP) and forty-eight options (on eight SERPs, reached by pressing a page number button at the bottom of the list).

Each session was completed when the participants had submitted the selected article that interested them the most from any results list provided. After submitting, they were asked to rate their satisfaction for the selected article, explain their strategy of choosing, and describe the usefulness of articles obtained from the synonym-search results list. After they had completed the final session, semi-structured interviews were conducted in a quiet office, all of which lasted between thirty and forty-five minutes.

### **5.3.1.2 PARTICIPANTS**

A total of twelve participants took part in this study, five males and seven females. They were all postgraduate students studying in different departments at the University of Bath, including: Computer Science, Chemical Engineering, Mathematics, Physics, Social Science and Mechanical Engineering. They were recruited via an online advertisement on the university noticeboard and their ages ranged between 25 and 40, with the average being 32.3. In terms of the self-reported level of expertise in the research field, one participant was a beginner, six considered themselves to be intermediate and five classified themselves as advanced. All reported using search engines for literature searches on a weekly basis. After completing the study, each was paid twenty pounds in cash for their participation.

### **5.3.1.3 MATERIALS**

The search application used was developed in PHP, a server-side scripting language for web development, in order to connect with the server-side synonym database. In addition, two lexical databases were integrated and used to retrieve synonyms: WordNet Thesaurus (Miller, 1995) and Moby Thesaurus (Ward, 1996). Moreover, a simple

stemming algorithm was developed to identify automatically the first noun in the keyword search term, which was then used to retrieve a synonym from the integrated database.

In addition, the keyword search term was stemmed by removing any question words (why, what, how, etc.), articles, and preposition words. Then, in turn based on an order of words in the remaining term, a part of speech of a word was retrieved from the database to check whether it was a noun.

A synonym was randomly selected in cases where more than one synonym was found, whereas an empty string was returned when none existed. HCI Browser (Capra, 2001), an open-source extension to Mozilla Firefox 3, was modified to provide the guidelines during the search task and to collect data through questionnaires at the end of each session. Figure 5.1 shows an example of a search interface presenting the keyword-search list and the synonym-search results list with two options while Figure 5.2 shows an example of an interface presenting the keyword-search list and the synonym-search results list with 48 options. All interactions were logged via screen recordings. Google was used to generate search results for both Keyword-search and Synonym-search lists. SERPs for each list were altered using Google's API.

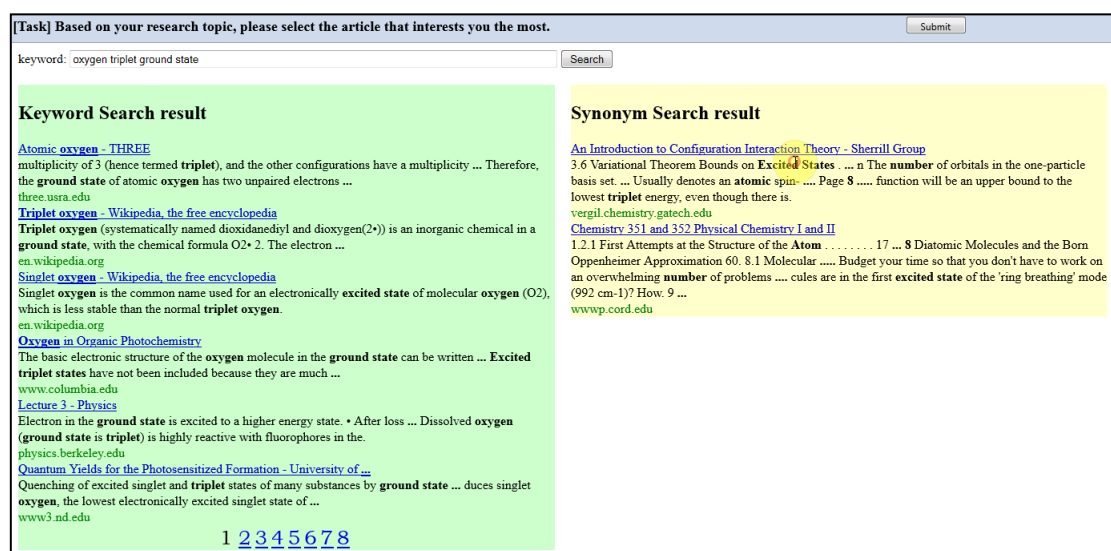
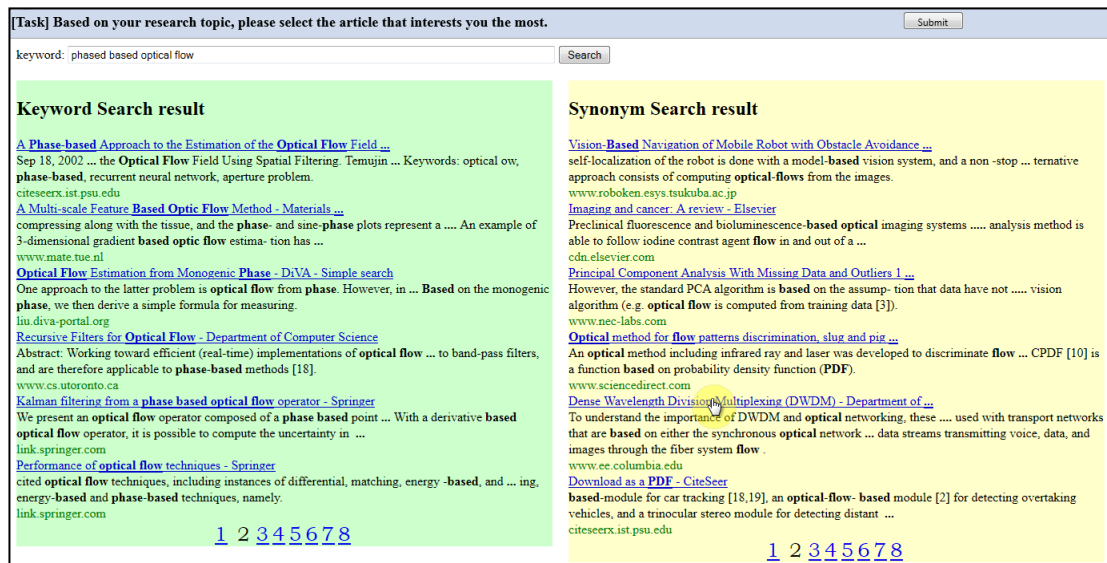


Figure 5.1 Study 5: An example of the computer screen presenting the keyword-search results and the synonym-search results lists with two options



**Figure 5.2 Study 5: An example of the computer screen presenting the keyword-search results and synonym-search results lists with 48 options**

After a selected article was submitted, the participants responded to the questionnaire mentioned above, which asked them about the purpose and topic of the search task they had completed, their level of satisfaction judgements, and their use of the synonym-search list. The survey protocol was as follows:

- What was the purpose and topic of today's search task? (Open-ended question)
- How satisfied are you with the search and selection process? (10-point Likert scale)
- How satisfied are you with your selected article? (10-point Likert scale)
- Was there any other article that interested you but you did not select it? Why did you not select it? (Open-ended question)
- Were the supplementary web pages from the synonym list useful? Why? (Open-ended question)
- How did you make use of the supplementary web pages from the synonym list? Please give an example. (Open-ended question)

In addition, the participants were allowed to take notes on paper during the search task, which they were able to take away with them after finishing the session. Before starting the next search session, they were asked whether they had performed any literature search outside the study on the same topic that they had searched for in the previous session. This was to elicit whether search activities performed in between the study

---

sessions had involved any chance encounters, i.e. serendipitous occurrences. However, only a few of them reported that they had engaged in extra searching and none of them declared that serendipity had happened.

The interview question schedule comprised open-ended questions that addressed the use of search engines for literature search purposes and the participants' experience with the provided application, as follows.

- What is your research area?
- How do you usually do literature searches? What do you like/dislike about search engines? What are the features or functions that you would like to have with your search engine?
- How did you find performing literature searches with this adapted search engine?
- What do you think about the supplementary search returns provided?  
Advantage/Disadvantage
- What do you like/dislike about this study?

The questionnaire to be used after each session and the interview question schedule were pretested in a pilot study of two samples in order to verify their appropriateness and comprehensiveness. No significant changes were made to the questions in either as a consequence of this study.

#### **5.3.1.4 PROCEDURE**

The participants completed each search task session individually in a laboratory, their office or in a quiet room. On the first day of the experiment, they were introduced to the HCI browser interface and informed that the general task in all five sessions would be to search for a scholar article related to their own current research topic that they had never read before. A sample search task was demonstrated to the participants, which used the phrase "the cause of UK flooding problems" as a search term and both search results lists were shown as well as being explained during this task. After each participant had confirmed that they understood the general task and the user interface, the first session started.

The participants continued the sessions on the following day, when they were informed that their search topic could be similar or different to the previous session, but still related to their current research. This process was repeated up until the last session. After

finishing the last session, the semi-structured interview was carried out. The interview was face to face, which was initiated by asking the participants background questions on their research area and how they generally undertook literature searches on the Web. They were then asked about their experiences when performing such searches using the application provided. The interviews ended with their being asked their opinions about the study design and the search task, which allowed for crosschecking whether they had convincingly performed the literature search tasks for the study as they would have performed in the real world context. In a debriefing session, the purpose of the experiment was explained to the participants, and they were allowed to ask any question related to the study.

### 5.3.1.5 ANALYTIC STRATEGY

All interviews were conducted, recorded and transcribed by the researcher, which were subsequently coded in order to allow her to become familiarised with the data at an early stage. Regarding which, open, axial, and selective coding, was used (Corbin & Strauss, 2008). Moreover, the screen recording and data collected via HCI Browser were used to complement the qualitative analysis. The analysis proceeded on a case-by-case basis, whereby each transcribed interview was completely assessed before moving on to the next. For this process, each transcribed interview was read several times and audit trails were constructed, which included an extract of the interview, a descriptive summary of each extract, and the initial theme. Comparisons between cases and preliminary interpretations were made after all twelve transcribed interviews had been analysed and the initial themes in each case identified. These themes were then clustered so that the final themes could be established.

## 5.3.2 ANALYSIS

### 5.3.2.1 DESCRIPTIVE AND BEHAVIOURAL RESULTS

#### 5.3.2.1.1 Iteration

Most of the search terms used in the study were phrases, for example, *multiple model adaptive control*, *wine fermentation analysis*, *child support in South Europe Latin America*. In addition, the participants sometimes entered a single word or a phrase that contained the unique terminology used in their research field. In these cases, synonyms of these unique words did not exist in the database and these were substituted by an

empty string. In the following examples the words that are underlined indicate those to be substituted and their synonyms:

[Participant #1] Search purpose: *“To find an article discussing how to detect and grasp symmetric objects.”*

Keyword search term: imitation multiple model

1<sup>st</sup> iteration: multiple model control imitation learning

2<sup>nd</sup> iteration: adaptive control imitation learning

Synonym search term: impersonation multiple model

1<sup>st</sup> iteration: full many model control imitation learning

2<sup>nd</sup> iteration: control imitation learning (Note: empty string returned)

[Participant #5] Search purpose: *“The keyword is “electron exchange energy transfer”. I would like to introduce this theory in my thesis so I need to read and understand it in order to use it as the basic theory in my thesis.”*

Keyword search term: Dexter transfer energy

1<sup>st</sup> iteration: energy transfer process

2<sup>nd</sup> iteration: electron exchange energy transfer

Synonym search term: starboard transfer energy

1<sup>st</sup> iteration: vim transfer process

2<sup>nd</sup> iteration: negatron exchange energy transfer

Each participant iterated their search terms in at least one of the five sessions, averaging out at 2.1 times per participant, which was mostly performed by adding or removing words in the search terms. However, in most of the sessions the participants used a single search term without iteration.

#### 5.3.2.1.2 SERPs viewed, position of the opened and selected options, and satisfaction rating

Dependent variables	Keyword-search list (48-option set)	Synonym-search list (2-option set)
SERPs viewed per iteration	3.1	1
Number of articles opened per session	5.1	0.4
Position of opened articles	6.9	1.5
Position of submitted articles	6.3	-
Satisfaction with submitted articles	7.8	-
Satisfaction with the search and selection process	7.5	-



**Figure 5.3 Study 5: Descriptive results under the 2-synonym-option set condition**

Figure 5.3 reports the mean number of SERPs that the participants viewed, mean number of articles opened and the mean ordinal position of opened articles on the keyword-search and the synonym-search list during the search process in the condition where two options were presented in the synonym-search list. The participants browsed through less than half of all the available SERPs in the keyword-search list, with average being 3.1 SERPs viewed per iteration for each search session (38.75%). The number of articles opened in each session was 5.1 (10.63% of the total of 48 options in the keyword-search list), which mostly were located within the first and the second SERPs. In addition, the average number of the synonym-search options opened was 0.4 per session (20% of the total of 2 options displayed in the list). However, all of the submitted articles under this condition were from the keyword-search list, with the average satisfaction regarding the submitted article and the search and selection process being 7.8 and 7.5, respectively.

<b>Dependent variables</b>	<b>Keyword-search list (48-option set)</b>	<b>Synonym-search list (48-option set)</b>
SERPs viewed per iteration	2.3	1.8
Number of articles opened per session	2.5	0.3
Position of opened articles	4.8	3.3
Position of submitted articles	6.8	3.0
Satisfaction with submitted articles	7.5	7.0
Satisfaction with the search and selection process	7.3	7.0

**Figure 5.4 Study 5: Descriptive results under the 48-synonym-option set condition**

Participants who were given the synonym-search list with forty-eight options browsed through several SERPs in both lists, with the average being 2.3 keyword-list SERPs viewed (28.75%) and 1.8 synonym-list SERPs viewed (22.5%) per iteration for each search session (Figure 5.4). Although those who were delivered a results list with forty-eight options did not report during the interviews that they were overwhelmed by the large number of supplementary results, the number of articles opened for both lists was less than that in the previous condition. That is, the number of the keyword-search options opened was 2.5 (5.21% of the total of 48 options) and the number of the synonym-search options opened was 0.3 (0.63% of the total of 48 options), with those opened being located towards the top of the lists. Most of the submitted articles were from the keyword-search list, except in the two sessions where the article from the synonym-search list was accepted. The participants' mean satisfaction levels with their

---

submitted articles chosen from the keyword-search and the synonym-search list were 7.5 and 7.0, respectively, whilst those for the search and selection process were 7.3 and 7.0, respectively.

#### **5.3.2.1.3 Synonym-search results selecting**

It is worth characterising the only two cases when the synonym-search results were submitted. In the first, the keyword search term was “the role of the mirror neuron system in motor” and the synonym search term was “character of the mirror neuron system in motor”. It appears that the meanings of both search terms were closed, yet not exactly similar. However, the substitution of the first noun found in the search term and its synonym kept the search term meaningful. This may be due to the fact that the selected first noun was not a unique word and its selected synonym had the meaning in the corresponding senses. This is similar to the second case, whereby the synonym search term preserved the meaning of the keyword search term. In this case, the definitions of the selected first noun and its synonym were quite similar, i.e. the keyword search term was “picture haze removal” and the synonym search term was “image haze removal”. This shows that the process of selecting the word to be substituted by its synonym is vital in the synonym-assisted search design in order to keep the search term meaningful to some extent.

#### **5.3.2.2 QUALITATIVE RESULTS**

To be able to clarify the results and make sense of participants’ experiences, themes were classified under two main categories: *literature search behaviours & requirements* and *experiences with the synonym search*. The first category encapsulates the participants’ literature search behaviour and the underlying requirements that influenced their behaviours. These requirements were elicited from the explanations of their use of features and functions provided by search engines. The second category captures the participants’ experience of using the synonym search and their interpretation of its efficacy. That is, it addresses our aim of investigating the benefits of the synonym search and how it can be employed in literature searches. The themes are described and discussed alongside direct quotations from the interviews so as to illustrate the rich data captured regarding the participants’ use of search engines in general and their views on the synonym search, in particular.

It is worth noting that participants were all research students who had some background knowledge on their research topic and they all reported that they found some interesting articles that related to their interest while undertaking literature searches for the study. In addition, the majority of the participants mentioned that working on the search task was quite similar to how they normally performed searches for their research.

[Participant #1] *“I found some articles that are interesting. It is fine to work on the task for five days because I need to do that research anyway. So for this particular experiment to spend time on searching a particular paper about a particular topic is something I need to do everyday. So I think it is useful.”*

During the interview some participants referred to the literature search tasks in this study as typical of their usual strategies in this respect, which suggests that the study design was quite naturalistic. However, some commented that sometimes they spent a longer time on literature searches in real world scenarios than they were expected to spend for current the study (thirty minutes):

[Participant #10] *“It was good in the way that I have to search for articles during this week, to spend about half an hour during the task. Well, sometimes we spend even more especially for the work. So it is nothing just being involved doing the task for half an hour everyday.”*

In addition, some participants who were provided a synonym-search results list with two options stated that they would briefly look at the supplementary web pages, but focus more on the keyword-search results list. For example:

[Participant #4] *“Sometimes they provided something good that maybe is related but it is not what I am looking for. So I just focused on the keyword search list and because it also provided me quite a few web pages. But for the supplementary, they just provided me two to three articles so I just look and if it is not relevant I just ignore it.”*

It seems that a large number of supplementary web pages being provided can engage participants in spending more time looking through the supplementary list, which may serve as evidence that search engine design should be modified to increase chance encountering. By being provided more options that to some extent relate to the search

---

term, users may be attracted to spending more time on browsing, which may result in their coming across unexpected useful results.

#### **5.3.2.2.1 Literature search behaviours & requirements**

##### **Gathering different variations of search results**

The research process typically involved a great deal of searching and gathering information related to particular subject matter. From the study, it was realised that various types of search results' content were required in order to consolidate the research concept. Some participants used search engines to gather specific information related to their research, such as that regarding the implementation of various techniques:

[Participant #6] *"I usually do it in Google Scholar, maybe because I'm in the second year and I still need some general information rather than the state of the art papers. Currently, I'm in the stage of implementing so I wanna know how people implement some stuff so I need more general rather than the latest news on articles so I generally use Google Scholar."*

In other cases, search engines were used to gather information in order to get a fundamental understanding of a research topic that was difficult to understand. A few participants mentioned that usually they had got the recommended articles from their supervisors that were hard to understand. In such cases, they used search engines to find materials that contained background knowledge and simply explained the main idea about the topics:

[Participant #4] *"Basically I've got something from my supervisor first about the recommended original paper list. I find some papers were quite difficult to follow and some might be very old. So I always start from using a normal one, like just Google it to look for some lecture notes or something that is quite simple to understand."*

The nature of some participants' work involved problem solving in programming language and they were mostly searching for articles that related to the technical aspects of this rather than theoretical work:

[Participant #7] *"I tend to use Google more, because I need information about practical stuff which I cannot get it in Google Scholar, more like technical not*

*theoretical... When you do programming when you sometimes have errors, it's like hidden errors, no one, you can't ask anyone, just people who're really doing this practical stuff about this technique, especially with low-level language, this kind of thing."*

In addition to the need to have the variation in search result contents, some participants required a variety of search result formats, such as image and video demonstrations, which explained the techniques used in the publications:

[Participant #10] *"I will just Google it. Put it into Google and maybe sometimes I will go to Google Scholar, because there is where I will get the papers or resources more about academic work. If I simply put the keyword in Google, then I probably get some commercial products, which just show me the result only without any explanation. Sometimes I may also search it by video because I am doing graphic work the video is very important. Some people may have some demonstrations... And also I would, because I am doing something in computer sciences, source codes are very important so if I get a paper that is just a paper, I would like to know what the related resources about the paper are."*

[Participant #5] *"I use both Google and Google Scholar. Normally when I use Google, because my keyword is quite specific, so then the first option that they gave is the Google Scholar link. Sometimes I click that then it's like access to normal Google Scholar. But sometimes if I just want some rough ideas, just the idea of what it is, I just use Google and Click Image immediately and see which kind of image I am looking for, just to give me an idea."*

### **Ranking algorithm reliability judgement**

The participants' explanations about their search behaviours showed that the judgement of ranking algorithm reliability played an important role in their literature search process. Most of the participants relied on the ranking algorithm of a particular search engine, for they seemed to believe that they would get valuable results from near the top of the results list. In addition, some participants mentioned that they only examined those search results located on the first results page:

[Participant #9] *"Well, sometimes if I have no idea which scholar I can follow, in that case I go to Google Scholar and then I just write my keyword. Then, I am*

---

*just kind of normally going for the first page, because they are more popular and we can see how many times they have been cited by other people. So I see whether they have been, you know, cited a lot or not. So, Yeah, Google Scholar is quite helpful.”*

Some participants stated that they had never used any refinement functions or known about search syntax that could facilitate their search process. This could be the consequence of over-reliance on the ranking algorithm that makes a user unaware of benefits they could get from using the search-refinement functions provided.

[Participant #6] *“Maybe there are those facilities, but they are not that easy to use and I’m not using them. Sometimes I search for something and it gives me or normally it usually comes with the other commonly used words. So sometimes I want to say that, for example, I want this word but not the other word in my articles. So maybe they don’t have that function or but if there is there maybe it is not very straight forward to use it.”*

However, other participants did not merely rely on the ranking algorithm and these cases the quality of search results was the major concern. The participants chose to use the search engine that employed the unreliable ranking algorithm, because the set of data source was of better quality. To compensate for the low efficient ranking, the refinement function was used to filter more relevant search results:

[Participant #12] *“Well, it’s horrible (the library search engine) but it is still quite good compared to Google Scholar if you spend some time getting rid of the links, the hits that are not related to your topic. I also use Google Scholar, both Google and Google Scholar, the problem with Google Scholar is sometimes it doesn’t return articles of, like very good quality or very good journals.”*

A few participants mentioned that sometimes they switched between different search engines that have different ways of ordering search results:

[Participant #8] *“For Google Scholar, compared to Microsoft Academic, I think they have more resources, I can find more papers... Because in Google Scholar I think they are only ranked by citations. But I think Microsoft Academic they have their own way to rank it even rank the people in this domain. I normally use*

*people ranking. I will search for several names on Google Scholar and then try to search the people in Microsoft Academic”*

The fact that some relied on the ranking algorithm may not be surprising. However, behaviour could affect chance encounters, the essential element in the research process. That is, the design of a search engine may somehow hinder user awareness of the possible useful results that may not be presented on the first results page.

### **Data source quality and search refinement**

Search engine users have different strategies to search and to refine their results to make them more relevant to their topic of interest. In this study, the participants’ literature search preferences reveal concerns about the suitability and reliability of the data sources. This concern seems to affect their behaviour regarding the choice of specific data source and the refining of a set of returns. Regarding which, some participants said that they started the search process by choosing a specific data source that provided specific results:

[Participant #3] *“Normally, most of the time I start in Google and Mm.. if it is a specific paper then normally it takes me to Google Scholar and that is how I continue the search. If I know the name of the author, I never even search in Google, I go to DBLP, which is the database for computer science. I go to that one directly and I search the whole papers of the author that I knew of... I hardly ever use the library website, because there are some papers that we can only access through the library website.”*

[Participant #5] *“Normally I do it in two ways. One is for something like quite a basic theory or like a background theory where normally I use Google and I use the keyword immediately... The second way is I use the Web of Knowledge, which currently it appeared to be Web of Science. That’s a bit more specific. It is really useful because I can just like do search for literature and refer back to a specific article by using Web of Knowledge.”*

Some participants applied search tactics, For example, using search syntax with the keyword, or using the search-refinement function to obtain more relevant results was one reported strategy:

---

[Participant #7] *“Sometimes I use (feature or functions), it depends, sometimes I look for reports, sometimes I look for tutorials, sometimes I want to be more specific I use specific type, or if I want to get the result for the specific year, for example in the last search, like 2013. I’m not looking for the very old one.”*

[Participant #12] *“Oh, I think, making sure to put more filters from the start can somehow improve the library site. So if I’m looking for articles, then I only obtain a list of related articles. Now regarding how to get rid of all these additional papers that are not related to my article, that are not related to field, I don’t know how they could do it but they should. Because I search for something and I get like the articles from like biology and physics, I don’t know why, maybe because it looks at the keywords and it returns a lot. And that didn’t happen in the past.”*

Based on the study context, that of searching for academic literature, several participants expressed their desire to have additional functions that would support their literature search. For example, one participant wanted a function that allows direct retrieval of reference articles:

[Participant #11] *“ Mm... for example as I said before, If I think a certain paper is useful for me, once I read it I will follow the references of the paper. I think if I just, for example, if I click one paper and if it provides the links of references or something like that, then it will be quite useful and much easier to find other papers I think.”*

Another participant mentioned about the ability to manipulate the ranking algorithm:

[Participant #7] *“Maybe they can add another button to order the result from the most recent one (by date), that’s helpful.”*

In some cases, users chose to use the search engine by considering the functions it provided. That is, these participants explicitly reported that the choice of search engine they used was mostly based on its supporting functions:

[Participant #9] *“And also about a function, this is a small thing, but if there is a function which goes to your EndNote, then I’d prefer using this one, like Google Scholar.”*



It seems that existing knowledge about data sources and experiences of carrying out online literature searches are necessary in order to identify the suitability and reliability of data sources. Hence, it would seem to be beneficial for researchers to be informed about the quality of literature sources and useful search functions.

### 5.3.2.2.2 Experiences with the synonym search

#### Search term and its synonym usage

Most of participants expressed their concern about synonyms of unique words, such as academic terms or the specific name of a scholar. That is, they pointed out that the search terms used in the literature search had to be precise and specific to the research topics. Consequently, the results obtained from searching by using synonyms were mostly irrelevant:

[Participant #1] *“For the synonym search, I do not find it very useful most of the time, because I think mostly it tried to find another word to replace my keyword and the other word it uses completely changes the topic I am looking for. I think it is because I am very specific in the thing I am looking for so the keyword I use, especially for academic search, the result is quite sensitive to the keyword because in academic work we need to be precise.”*

Some participants mentioned that the results from a synonym were more relevant to the key topic when they entered search terms that had a general meaning:

[Participant #5] *“In my point of view, if you search using a keyword that’s not too scientific and quite a general background, it seems, like, the synonym search result mixed quite a variety of supplementary web pages, rather than go straight to the exact result. They tried to find something useful, so I think basically it depends on the keyword if your keyword is general enough, it probably gives really useful options for you.”*

[Participant #3] *“About the synonym, to be honest, it was not of any help. I think I can understand the reason. So let’s say when I search for something on Google, let’s say, the other day for example, I was searching for ‘how to remove a stain from my carpet’, So I said I got some results and then I tried with ‘how to clean my carpet’ then I got it. But when it comes to the technical work, for example, I was looking for ‘answer set solvers’ this is the technical word, so we call them*

---

*solvers and there are different sets of solvers. So the synonym one they changed 'solvers' to 'calculator', the phrase was 'answer set calculator' I imagined. And then it obviously does not work in that context. We do not have such a technical term. So it came up with some stuff that was not relevant at all."*

Several participants suggested that providing guidelines of how synonyms were selected would be helpful for submitting the appropriate search term and judging the relevance of search results. In addition, a list of synonym words could be provided to guide the user to search for the most appropriate words:

[Participant #8] *"Sometimes, for example, if I want to search for coffee, there are many kinds of coffee but I don't know like the name of like cappuccino or latte. Probably I don't know the terms of these, I only know coffee. So probably if I search for coffee, if it can come up with these keywords then this is what I want. I think it will be good. So probably if it can list these things up, it will be useful."*

[Participant #5] *"I think if you want to make the supplementary web page really useful, probably you should have like help functions to suggest what kind of keyword, meaning, or what kind of mixing that the supplementary web page will give you or will generate for you."*

### **Variety of synonym-search results**

Several participants stated that the results from the synonym search were, to some extent, relevant to their search topic:

[Participant #4] *"To compare both, I do not think the supplementary web page is useful for me because some words in mathematics are quite unique words. For example, some words like 'mixture distribution' in the context of statistics, but the synonym search just provided me some irrelevant topics that are nothing related to mathematics at all. But sometimes they provided something good that maybe it is related but it is not what I am looking for."*

A few participants mentioned that searching by using a synonym sometimes returned useful results that indirectly related to the main search purpose, such as general ideas or useful background about the search terms:

[Participant #5] *"They gave not exactly what I am looking for, but something that was quite basic and related to something I'm looking for. So I have a look, it is quite useful but even so I did not submit it because it is not based on what I'm interested in."*

Some participants explained their experience of finding some relevant results from other research disciplines:

[Participant #3] *"For example, as I told you that the approach that I am using is argumentation, it has its root from psychology. Once they take that from psychology or philosophy or whatever when they come to computer science or other disciplines normally the terminology can change a bit. When you replace it often I got some results from psychology departments or some other disciplines, which I can't say are not related. They might be sharing something but it was not specifically what I was looking for, because they are different in my discipline. It was interesting in some ways to see the original paper from a philosophy department and so on that has come up with the idea first and then later on many people adapted it, many people from computer science. So it was interesting but it wasn't what I'm looking for. So I would say it is kind of metadata, maybe it is good in that sense."*

[Participant #4] *"It might be useful when, maybe when you try to find something that's in the other areas. For example, what I'm looking for is something in mathematics or statistics, but maybe in biology or chemistry scientists they also use this kind of knowledge or keyword as well. Maybe I can have a look at the content in different areas later. Do you know what I mean? Because right now I just focus on my research, I just want to know what exactly the explanation about the keyword is. But maybe some people in another area they also use this thing as well. And if this shows maybe I can have a look later. I mean if it shows in the supplementary, maybe if I'm lucky it will just link to some things that are related, maybe it's some article that mentioned this topic in some other areas."*

However, other participants reported coming across unexpectedly useful results returned from the synonym search:

---

[Participant #10] *“For example, I remembered that one time I wanted to search for the image de-hazing algorithm. Yeah, from the keyword search I can get the paper and the paper does the thing but I don’t know that if I could get the depth information for de-hazing the image as well. So I found it, I actually found the paper from the additional, the synonym search, on there. Although it is not the thing I am searching for but that thing really interested me, because I did not know I can get the depth from that, I don’t know, it is like a surprise for me.”*

[Participant #12] *“I mean it could it expand your search and you could get a website that you are not thinking of. So it’s almost like doing two searches simultaneously.”*

[Participant #1] *“I think once the synonym search found something relevant that I did not find in the key word list. So I submitted that result from the synonym search but I can’t remember in what circumstance exactly.”*

Based on the results, it seems that the synonym search somehow provides a variety of relevant search results. That is, on the synonym-search results list, the participants came across some general background information, related research from other disciplines and some unexpected useful results.

### **Relevance of synonym-search results**

Although several participants suggested that the synonym search could sometimes return useful results, others reported that most of the results in the list were not relevant to the topic area. Hence, the results from the synonym search could be more useful if more relevant results were retrieved.

[Participant #6] *“ To say how those results are comparable. I mean usually I’m not in favour of synonym search because sometimes it’s misleading, coz when it found it synonyms sometimes its synonym is not really related to what you’re doing.”*

[Participant #7] *“For the synonym list, the result is a little bit related, but most of them are not that good. I don’t know, but when I, for example, when I wrote some keywords about ‘pipeline multicore’, I used the plus sign between pipeline and multicore. In normal search they gave me result that only related to*

*computer science, but with synonym they just gave me information about gas, about factory, about how you can you use a pipeline in a factory, which is a completely different, different topic.”*

[Participant #10] *“The main benefit is I can get additional information. The main drawback is it sometimes it gave me too much irrelevant information. For example, when I am trying to search for an algorithm, and then it just gave me something about an unrelated algorithm. For example, I wanted to search for, ... I remembered I wanted to search for shadow detection algorithm, and then the additional result showed me something about a sorting algorithm, general sorting algorithm. Yeah, it is totally not the thing I want.”*

### 5.3.3 DISCUSSION

The present study sought to investigate literature search behaviour in real-world circumstances and to explore the potential usefulness of automated synonym-enhanced searches. The aim was to gain insights into how search engines could facilitate literature searches and enhance serendipitous experience. Although the work mainly focused on the literature search context, other dimensions related to information seeking during the research process were also probed.

The uses of search engines for academic purposes were found to be different based on the information needs in distinct research areas. The requirements for obtaining variety in search results would appear to cover two dimensions: variety content and in types of results materials. These findings suggest that search engines should aim to support researchers by presenting search results from different research aspects, and with various types of materials. With this implication of search engine design, researchers would be more attentive to the connections within their research field or between different areas of research, which may increase the chances of serendipitous encounters.

Existing knowledge about information sources, search functions, and search tactics would appear to affect literature search ability with regards to the extent to which the targeted results are retrieved. That is, the outcomes have shown that those participants who had knowledge about information sources could directly identify where the results of interest were likely to be located. In addition, the use of search functions and tactics can improve the overall relevance of search returns. In order to increase search ability,

---

researchers' awareness about information seeking environments should be raised in several aspects. First, information about the suitability of data sources for a particular research area should be made available. Second, useful search functions and tactics should be promoted. Moreover, it emerged that the participants wanted the ability to manipulate the ranking of search returns. Hence, it is contended here that rather than filtering relevant search results, the facility of changing the ordering of search returns based on the searchers' preferences may be a more suitable way of supporting their information needs.

The participants' general perceptions on the use of synonyms in information seeking were mostly positive. Some reported that the synonym search was useful in terms of the way that results from different aspects were presented, although often these were not what they were targeting. However, several participants held the contrasting point of view that the results were misleading or irrelevant. Although the perceived relevance of search results may partly be based on the subjective judgement, this irrelevance perception can be put down to the low efficacy of the synonym search, to some extent. It was pointed out that some words should not be substituted by synonyms, whereas the substitutions of other words could be useful in order to retrieve unknown relevant results. This seems to point to an interesting implication for improving the design. That is, perhaps, a search engine design that allows users to mark which word in the search term is worth exploring in terms of synonyms, rather than just the first noun is the current case, this may help them expand the boundary of their search topics and retrieve more useful search results.

Regarding serendipitous experiences, the findings revealed that the results from the synonym search could occasionally increase the chance of encountering unexpected information through the retrieval of a wider variety of search results. Additionally, synonym-search results sometimes presented connections with different research disciplines, which suggests that this facility could be integrated into search engines in order to enhance the chance of encountering serendipity. Furthermore, it emerged that the number of supplementary web pages affected the search behaviour by engaging searchers to browse deeper through the supplementary list.

In summary, it is worth stressing that this is early exploration of a general idea, and with a rather unsophisticated implementation. As discussed above, the whole idea of

“designing” for “serendipity” is somewhat paradoxical, which is why it was thought worth exploring a simple implementation that has a high chance of yielding some unanticipated search returns.

### **5.3.4 STUDY LIMITATIONS AND FUTURE WORK**

Although the participants who took part in the study were from different research disciplines, it was not possible to identify clearly the differences in literature search behaviours and requirements across these. Hence, it is proposed that the behaviours of a larger number of participants from each discipline be investigated in order to provide a deeper understanding of these different support needs.

Regarding the synonym search, the quality of the supplementary web pages from the search may have been limited in two ways. First, a simple stemming algorithm was used to identify a single word to be substituted by its synonym. A better algorithm could be employed in such a way that its synonym exists and the substitution adds value to the set of supplementary results. Second, the synonym database used in this study was not developed specifically for academic terminology. In particular, a number of synonyms in the database were not used in literature, which consequently, yielded no results or unrelated ones in the supplementary lists.

## **5.4 CHAPTER SUMMARY**

This chapter has presented a qualitative study of information seeking behaviour in the context of scholarly literature search. The insights obtained into the literature search behaviours and requirements provide better understanding of how search engines are used to support the research process in real-world circumstances. Instead of investigating the instances when individuals come across serendipity as in past research, the focus was on the users’ underlying requirements, with the aim being to elicit the implications for improving search engine designs to increase chance encountering. Furthermore, the study conducted in this chapter has extended the earlier finding reported in Chapter 4, which indicated the potential usefulness of synonym searching. That is, the usefulness of automated synonym-enhanced search was explored in order to identify potential search designs for enhancing serendipity encountering. In the context of literature searching, this work has revealed that the synonym search might be useful in order to

---

provide a variety of search results, which may increase researchers' opportunity to come across serendipity. That is, the research involved taking one step towards designing for a serendipitous experience by implementing a model that can be integrated into search engines, the benefits of which were provided through real-world literature search tasks.





## **CHAPTER 6**

### **CONCLUSION**

#### **6.1 SUMMARY OVERVIEW**

The principle objective of this thesis has been to investigate the effect of the number of search results and their ordering on search satisfaction and behaviour. The overarching aim has been to understand the determinants of user satisfaction in order to provide implications for the improvement of search user interfaces, provide some underpinnings for user-centred design. In this chapter, the work presented throughout the thesis is summarised and discussed as follows. Firstly, the findings from the empirical studies are reviewed regarding the four main research questions. In addition, some possible connections between these empirical results are drawn together. Having described an integration of the findings from the experimental work, the discussion then turns to consider how search satisfaction may operate more generally and thus, a theoretical sketch of search satisfaction judgements is presented. Next, the thesis contributions are discussed in relation to the main research questions followed by the limitations and future directions.

#### **6.2 THESIS SUMMARY**

##### **6.2.1 A REVIEW OF THE THESIS FINDINGS**

To begin with, a literature review was conducted to examine the factors that could influence satisfaction judgement during a decision process, which allowed for the identification of the search engine design features that potentially affect search

satisfaction. It was found that too many options to choose from can lead to lower satisfaction with the decision outcomes in search engine use as well as other contexts. The first research question (**RQ1: Does choice overload affect users' satisfaction with search engine use?**) was devised with the aim being to explore the choice overload effect in an online environment that closely modelled search engine use in a real-world situation. In particular, effects stemming from the extensive number of search results returned were investigated through open-ended search tasks.

The first experiment in Chapter 3 provided no evidence that choice overload would affect users' satisfaction regarding their selected search results. Indeed, the effect of the number of search results worked in the opposite direction, with the tendency (although only marginally statistically significant) being more satisfaction reported by the participants who chose web pages from the large set of returned snippets, rather than those under the small set condition. This result contrasts with the findings of a previous study (Oulasvirta et al., 2009) that reported a negative effect from a large set of search returns. In response to the outcome from Experiment 1, Experiment 2 was conducted to investigate this issue further as well as to explore the effect of the ordering of returns on users' behaviour and satisfaction, as these seemed like a major potential explanation for why choice overload would not occur. This second experiment addressed the second research question (**RQ2: What are the effects of the ordering of search results on the behaviour and satisfaction of search engine users?**). The findings from the second experiment provided no evidence that a set size of options would affect user satisfaction with the selected options. However, the participants' behaviour revealed that they were influenced by what they believed to be true about the ordering of returns in such a way that they would select options displayed early on in the search results list when they perceived the ranking was reliable. More specifically, the participants chose from further down the set of returns when they believed the ranking was random, and when more returns were available (i.e. the Set Size was large). To model this behavioural effect with a Bayesian updating of link-probability used in the SNIF-ACT 2.0 model would require a model in which prior probabilities decreased according to some pre-specified function through the list of returns. It seems clear that participants are predicting the likely variation in quality among yet unseen members of the choice set.

In addition, the participants rarely committed to select the chosen web pages when they were first encountered and instead, they tended to browse through more options, but not

---

all of them, before eventually returning to select the chosen ones. This was supported by the findings whereby although the number of SERPs viewed in the large set condition was mostly beyond the first SERP, the majority of the selected web pages were located in the first results page. This behaviour is like that discovered in menu-search tasks (Brumby & Howes, 2008), and can readily be understood in terms of the models of Cox and Young (2004) and/or SNIF-ACT (Fu & Pirolli, 2007). It demonstrates, therefore, that key elements of these models do extend to this somewhat different decision task.

The third experiment was then conducted to explore whether time pressure is a moderator of the choice overload effect in this context. The findings from this experiment showed that when time was more strongly constrained, choices from small sets led to relatively more satisfaction. In other words, the relative satisfaction with web pages chosen from larger as compared with smaller sets interacted significantly with the time available to choose, which revealed the important role of time pressure as a determinant of choice overload. This particular finding is the main finding in the thesis to suggest that, in certain contexts, choice overload may indeed be an issue for search engine users. But further, it shows that it is a subtle effect, very dependent on the time pressure the searcher is working under.

Furthermore, the results confirmed that the ordering of search results influenced users' selection behaviour and satisfaction regarding the selected web pages in that the participants chose web pages from near the top of the list in the reliable ranking condition, but selected the web pages from further down the list when the options were randomly ordered. The use of a search strategy was also influenced by time pressure, such that when this was strongly constrained, the participants were likely to make their decision based on the level of familiarity of the web sites and/or relying on the ordering of options when they regarded the ranking algorithm as being trustworthy.

Experiment 4 followed up the previous investigations on the choice overload effect by increasing the number of options in the large set of search results returned. Some additional variety among the search results was introduced by providing supplementary web pages for each task in the separated lists, in order to explore users' reliance on the ordering of returns given this extra feature in two task frames: seeking a single target or three targets. The aim was to investigate whether this introduction of variety would result in the participants searching through more SERPs in the main results list and

hence, addressed the third research question (**RQ3: Does a search task which encourages users to seek various targets affect users' behaviour and satisfaction in search engine use?**). The experiment was also designed to investigate the relation between satisfaction with the selected option and with the search and selection process, by simply asking participants to make separate satisfaction judgements for each of these. In addition, the satisfaction on the selected option was measured according to two aspects: the first is that the selected option is judged based on its value to the current task, while the second pertains to a judgement of satisfaction independent of the task. The findings confirmed that searchers were not typically affected by the choice overload effect, in that the tendency was for higher satisfaction being reported from the participants who chose from the larger set of options. Furthermore, the results revealed that in the situation when a single web page was to be selected they mostly overlooked the supplementary web pages displayed in a separated list, preferring to choose options presented in the ordinary ordered list. However, in the circumstance that multiple options were to be selected, the participants' behaviour revealed their need for variety among the search results. In addition, the supplementary web pages generated by a synonym search were selected more than those generated by a quartile search and the search and selection process was reported as being more satisfactory when supplementary web pages from the synonym search were presented. Finally, users' satisfaction judgements on the search and selection process were strongly correlated with that regarding the selected options, which implies that the satisfaction with the search process could influence that for search outcomes.

Building on the findings from Experiment 4, the fourth research question was formulated (**RQ4: How can an automated synonym-enhanced search support a literature search process and increase the chance of encountering a serendipitous experience?**). Based on the usefulness from the synonym search reported in Experiment 4, an automated synonym-enhanced search was developed and evaluated in Study 5. While the earlier experiments were focused on measuring search satisfaction based on a set of search tasks, this final study investigated literature search behaviour in real-world circumstances through a longitudinal study and semi-structured interviews with research students. The aim was to explore the potential usefulness of automated synonym-enhanced searches and gain insights into how search engines could facilitate literature searches by enhancing serendipitous experience. The results revealed that the

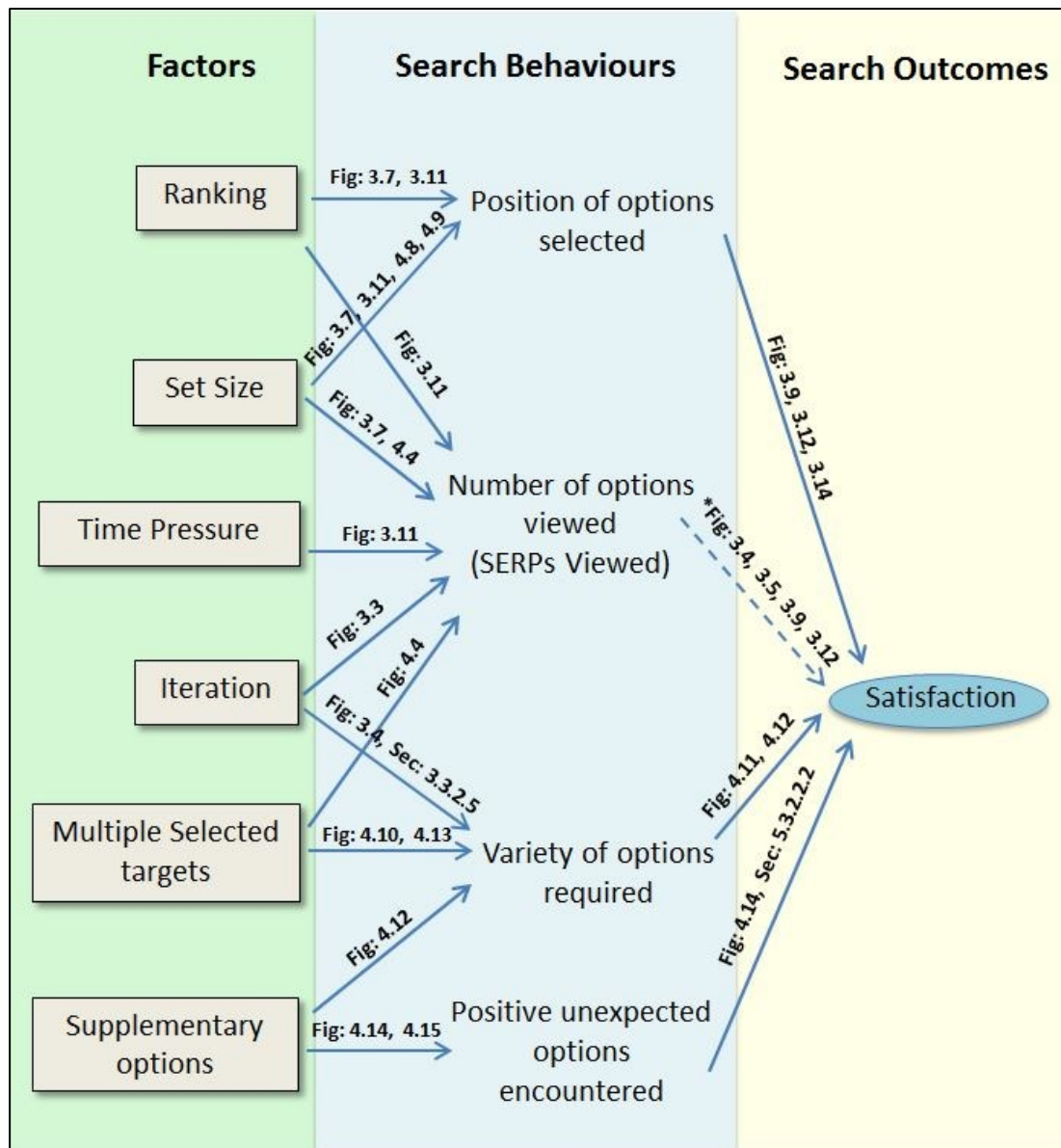
---

participants' general perceptions on the use of synonyms in information seeking were generally quite positive and suggested that a synonym search might be useful for providing a variety of search results, which consequently increased the researchers' opportunity to come across serendipity. Some suggestions for developing the simple implementation of the idea were apparent from participants' comments.

### **6.2.2 AN INTEGRATION OF THE THESIS FINDINGS**

Figure 6.1 summarises the effects of the factors manipulated in all of the studies presented in this thesis on search behaviours and outcomes. The figure indexes the findings in the thesis in terms of the figures and subsections in the empirical chapters which display each significant effect. Firstly, the empirical results revealed that there are at least two factors that affects user behaviour of selecting options including a ranking and a set size of search returns. More specifically, a highly-reliable ranking of search results returns influenced users to select options that were presented towards the top of the returned list. However, when the larger sets were presented, the positions of the selected options were located further down the results list. Secondly, the perception of ranking reliability affected the number of options to be assessed during the search process, such that more options were browsed when a ranking was perceived to be less reliable, i.e. randomly ordered.

In addition to the effect of ranking, the number of options viewed and/or SERPs viewed was affected by the set size of available options, time pressure during the search process, the number of options to be selected as well as iteration behaviour. Next, a variety of options was required when users were allowed to iterate their search terms, select multiple choices, and/or when supplementary options were provided. Furthermore, users' experience of encountering unexpected useful options occurred when supplementary lists generated by a synonym search were provided. Finally, their satisfaction judgement was influenced by the search process regarding those search behaviours identified, as shown in Figure 6.1. In this regard, it is worth noting that satisfaction judgement could be based on different aspects, such as satisfaction with a selected option or satisfaction with a search and selection process. In addition, when the desire for variety was perceived as having been fulfilled, the users gained confidence in their decision, which in turn led to search satisfaction.



**Figure 6.1: The effects of the factors manipulated in the thesis studies on search behaviour and outcomes**

In Figure 6.1, the dotted arrow (\*) shows the possibility that the number of options and/or SERPs viewed may lead to satisfaction, however the main effect of this was marginal at best in the thesis experiments.

### 6.2.3 A THEORETICAL SKETCH OF SEARCH SATISFACTION JUDGEMENTS

Regarding the findings presented, an overall guiding sketch of the cognitive and behavioural processes can be attempted, which unpacks the search process a user would follow when performing a search and subsequently evaluating the search outcomes in a

---

search engine use context. The rationale for the model is based on the thesis findings as well as understanding of search behaviours in rather different search task contexts, e.g. choosing one correct item from an ordered list of items in a menu (Cox & Young, 2004; Brumby & Howes, 2008; Fu & Pirolli, 2007). Although this process modelling was not a main objective of this thesis, by drawing together the existing theoretical models with the current findings on the subject of how people search and make judgements about search satisfaction, new theoretical insights have been provided to this field. The stages of the model are described as follow:

The first stage, *search and selection*, involves the cognitive process of (presumably) defining a threshold value above which an option will be considered for selection based on the judged level of relevance of this option for reaching a search goal. The findings from Experiment 2 and 3 revealed that the belief about reliability of the ordering of options may influence the establishment of threshold in this phase, such that a threshold value of a targeted option may be based on the value of items presented earlier on the results list when the ordering is believed to be reliable. However, when the ordering of options was perceived to be random, a threshold value may be sensitive to context, that is, it may be adjusted based on the options assessed so far during search process. In the next phase, options could be assessed individually or as a set in order to perform further actions. That is, when the estimated relevance of an entire set is lower than the defined threshold, users would iterate or change their search terms in order to retrieve a new set of options (e.g., in Experiment 1). On the other hand, the estimated relevance of individual options can affect users' further search actions, whereby they may choose to assess more options, select the current option or return to a previous option to reassess or select it. In addition, the findings from Experiment 3 and 4 showed that the further action performed also depends on the search process constraints, such as time pressure, the number of available options and their variety.

In the second stage of the model, *satisfaction with the search outcome* is judged based on its content, especially in relation to the search task and perhaps also in relation to the judgement of the snippet that prompted this item to be selected, and presumably combined with a broader judgement of the efficiency or completeness or ease of the search strategy employed during the search process. This is derived from evidence reported especially from two experiments, the effect of time pressure in Experiment 3 and the use of different satisfaction scales in Experiment 4, where that satisfaction with



the search and selection process was strongly correlated with judgement of satisfaction in the selected options.

The proposed model may help scholars in future with regards to the steps employed in search engine use and their exploration of the factors that affect search behaviour at each step whilst at the same time measuring user satisfaction. Furthermore, the model could be guidelines for system designers so as to enable them to support better each step of a search and selection process.

### **6.2 DISCUSSION OF THE FINDINGS AND CONTRIBUTIONS**

In this section, the findings from the studies presented in the empirical chapters and their contributions are discussed in relation to the four main research questions.

#### **RQ1: Does choice overload affect users' satisfaction with search engine use?**

##### **Contribution 1: Extending the current understanding of the choice overload effect in the context of search engine use**

Based on the primary research question, the first contribution of this thesis is to extend the current understanding of choice overload in search engine use in three respects. Firstly, the effect in an online environment that closely reflected real search-engine use was explored, which is novel to the literature in this context. Secondly, the findings from the experiments have shown that choice overload does not typically affect search engine users, because a large number of search returns emerged as being associated with a positive judgement regarding the search outcome. In spite of these results being in contrast to those of previous studies conducted by other researchers, such a positive effect was found throughout the series of experiments. Furthermore, the number of options was expanded so as to be greater than those ever investigated by previous research in this context. Moreover, this work has demonstrated that perception from a search process influences the satisfaction judgement with the search outcome. In this sense, the findings support the process-product leakage theory of choice overload (Scheibehenne et al., 2009) in a context of search engine use.

#### **RQ2: What are the effects of the ordering of search results on the behaviour and satisfaction of search engine users?**

---

**Contribution 2: Identifying the effects of ordering of search results on user satisfaction regarding the search processes and outcomes**

Prior studies have shown that the ordering of search results influences search engine users' behaviour during the searching and selecting processes, such that mostly only search returns on the first SERP are selected. Despite the findings from this thesis having revealed such search behaviour (Experiment 1, 2 and 3), the investigations were extended to the understanding of this phenomenon by exploring the effect of such ordering on the users' satisfaction with the search outcomes. Consequently, the second contribution of the work presented in this thesis is that it provides evidence that the perceived reliability of the ordering of returns influences search behaviour and satisfaction judgement with the selected options, thereby warding off any effect of choice overload. That is, when the ranking is perceived as reliable, the considered options are mostly located on the first search results page and the selected web pages tend to be chosen from near the top of the result lists. However, users will browse deeper through the list and selected web pages from further down, if they believe that the ordering lacks reliability, i.e. when there is random ordering. Therefore it is concluded that the reliable ordering of options can affect a relevance judgement of options in this context, such that the users may be biased towards assessing options ranked higher in the list as they perceive these options as being of greater relevance and hence creating higher satisfaction. In addition, the participants' search behaviour showed that they were less likely to stop searching once the value of the most recent option considered had reached their subjectively defined threshold. That is, they tended to continue browsing more options, even though the selected item was found early on the ordered list of returns, and they eventually returned to select the chosen item. This suggests that under this arrangement more options are assessed in order for the users to increase confidence in their decision and to reduce the risk of neglecting any potential high-value options that are located further down the list, however, it would appear that not all available options will be assessed.

**RQ3: Does a search task which encourages users to seek various targets affect users' behaviour and satisfaction in search engine use?**

**Contribution 3: Extending the current understanding of variety seeking in the context of search engine use**

In terms of users' behaviour, the ordering of search results led to an adaptation of their search strategy in order to seek for variety from the list, rather than relying on the ordering provided. That is, in the situation where multiple options were to be selected, users tended to opt for supplementary web pages or chose some options from further down the ranked lists, which goes against the expectation that items from further down the ordered list would be less relevant to the task. Moreover, this thesis contributes to the current understanding of variety seeking in search engine use by demonstrating that a synonym search can increase variety among search results, which in turn leads to more satisfaction with the search and selection process. Further, the findings have revealed that supplementary web pages are more likely to be selected when a large set of returns is provided rather than a small one, which may be evidence of a choice overload effect in that the users avoided going through a long results list of search returned.

**RQ4: How can an automated synonym-enhanced search support a literature search process and increase the chance of encountering a serendipitous experience?**

**Contribution 4: An examination of the usefulness of an automated synonym-enhanced search to support a literature search**

The issue about chance encountering and serendipitous experience in online information seeking has been investigated and discussed in many search contexts (e.g., André et al., 2009b; Foster & Ford, 2003; McCay-Peet & Toms, 2011). The previous research has either tried to identify the activities engaged in at the time of encountering serendipity and the types of unexpected information retrieved or proposed models for increasing chance encountering, but they have rarely attempted to investigate the users' satisfaction when using those models in real-world circumstances of literature searches. The final contribution of this thesis is that of having modelled an automated synonym-enhanced search and illustrated how well this may increase chance encountering and lead users to come across serendipitous experience in the real-world task of literature searching. The findings from the study have provided evidence that a synonym search might occasionally be useful in order to provide a variety of search results including unexpected ones, which in turn may increase researchers' opportunity to come across serendipity. Since a synonym search has not been used in such a way for previous research, this work contributes towards the designing of a serendipitous experience by

---

implementing a model that can be integrated into search engines, the benefits of which to users have been shown through real-world literature search tasks.

### **6.3 LIMITATIONS AND FUTURE DIRECTIONS**

Throughout this thesis the limitations of the work have been discussed within the respective chapters. This section discusses the other limitations of this thesis in wider perspectives and proposes directions for future work.

#### **Sampling and Generalisability**

Regarding the wide range of the general population of search engine users, one of the main limitations of the research presented in this thesis is the relatively small sample size for each experiment and study. In addition, although all the participants reported that they used search engines in their everyday lives, most were undergraduate and research students who were in academia. Therefore, the findings from this work cannot be representative of other populations in other communities.

In addition, in this work the different levels of expertise when using search engines was not a main focus of the investigations. However, it was anticipated that individual differences in searching skills could affect the users' behaviour and satisfaction in this context. In particular, the difficulty from the search process caused by the presence of a large number of options could be more severe for those who lack these skills when compared with experts who often use search tactics and functions to discard the irrelevant returns (Hargittai & Hinnant, 2008; Aula et al., 2005). Further work may wish to examine the effect of levels of users' search expertise on their search behaviour and satisfaction. With such understanding, the design of search engines can be improved by providing better support for beginners who could feel frustrated or discouraged by the online search process.

Furthermore, this work has only explored the use of search engines on a desktop computer. With other electronic devices, such as mobile phones or tablets, the use of search engines could differ according to the purpose and situation as well as being limited due to the smaller dimensions of the display screen. The presentation of search results could possibly negatively affect the search process, for example, in the situation where the user needs to find information rapidly on a mobile device and receiving a large

set of returns may complicate the process, thus affecting their decision behaviour and satisfaction.

### **The Effect of Variety of Search Results**

Although the qualitative findings from the literature search study in Chapter 5 reported the users' requirements of a variety of search results from different aspects, the effect of this on users' satisfaction was not directly investigated. It was anticipated that even though users can gain advantages from the presentation of search results from various aspects, variety introduced by too many aspects might well increase the complexity in the search process, which in turn may cause users suffering from the choice overload effect in this extent. Future research could be carried out to investigate the effect of variety of search results in this online information-seeking context for a deeper understanding of how variety among search results should be presented.

### **Automated Synonym-enhanced Search**

Despite the limitations of the synonym selected algorithm and the synonym data source mentioned in Chapter 5, another limitation of the study is that the automated synonym-enhanced search was integrated with Google web search. For, as the participants reported during the interviews about their use of different search engines for their literature search activities, the investigation might not be entirely realistic in this case even though the literature search tasks were based on each participant's actual research topic. That is, due to the fact that a single search engine was used in the study, this might affect the perception of the relevance of search results generated by the synonym search. If this is the case, then the model of the synonym search should be further developed to be able to integrated with other academic databases and search engines used in literature search activities, e.g. Web of Science, ACM Digital Library, etc.

An additional limitation is that the usefulness of the model was investigated in the specific context of a literature search. Based on the wide use of search engines, the efficacy of synonym search model could be explored in other contexts, such as information searching in real-world everyday life circumstances. Moreover, instead of using a pre-defined search task, the user could be allowed to perform the search based on their own selected topic of interest.

---

### **Satisfaction questionnaire**

It is worth noting that satisfaction ratings for all of our studies presented in this thesis were measured through 10-point Likert-type scales with positively worded questions and satisfaction increasing from 1 to 10. This may introduce biases and affect the participants' responses to some extent. For example, the participants may tend to agree with all or almost all statements in a questionnaire regardless of each question's content (acquiescence bias) or they may respond with extreme rating scores rather than select points near the middle of the scale (extreme bias). Questionnaires with a mix of both positive and negative wording are commonly employed in order to reduce these biases (Anderson et al., 1983). However, more recently it was found that alternating question items could also be problematic, such that this could lead to users' misinterpretation and mistakes as well as researchers' miscoding (Barnette, 2000). Sauro and Lewis (2011) conducted two experiments to investigate the effects of questionnaires' wording, the alternating question compared with all-positive versions. The findings revealed no benefits from including both positive and negative items in the questionnaire compensated the disadvantages of mistakes and miscoding. In addition, the scores from both questionnaire versions were similar. Therefore, the researchers suggested that an all-positive worded questionnaire could be used with confidence due to the fact that respondents are less likely to make mistakes when rating and results are less likely to be miscoded.

## **6.4 CONCLUDING REMARKS**

This thesis has advanced the field by extending understanding of the effects of the number of search results returned by a search engine and their ordering on users' behaviour and satisfaction. In particular, the work has involved exploring three specific issues, namely, choice overload, variety seeking, and serendipity. To this end, a series of studies that employed both quantitative and qualitative approaches were used to investigate these effects as well as to illustrate that the implications of design proposed can be applied in order to enhance the system designs to support the users' needs. The work presented in this thesis takes one step towards a more complete understanding of the effects of search engine design on search satisfaction and behaviour.



## REFERENCES

- About WordNet*. (2010). (Princeton University) Retrieved August 1, 2013, from WordNet: <http://wordnetweb.princeton.edu/>
- Agichtein, E., Brill, E., & Dumais, S. (2006). Improving web search ranking by incorporating user behavior information. *In Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 19-26). ACM.
- Anderson, A. B., Basilevsky, A., & Hum, D. P. (1983). Measurement: Theory and techniques. In P. H. Rossi, J. D. Writ, & A. B. Anderson, *Handbook of Survey Research* (pp. 231-187). New York: NY: Academic Press.
- Anderson, J. R., Matessa, M., & Lebiere, C. (1997). ACT-R: A theory of higher level cognition and its relation to visual attention. *Human-Computer Interaction*, 12(4), 439-462.
- André, P., Teevan, J., & Dumais, S. T. (2009a). From x-rays to silly putty via Uranus: serendipity and its role in web search. *In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 2033-2036). ACM.
- André, P., Teevan, J., & Dumais, S. T. (2009b). Discovery is never by chance: designing for (un) serendipity. *In Proceedings of the seventh ACM conference on Creativity and cognition* (pp. 305-314). ACM.
- Apted, S. M. (1971). General purposive browsing. *Library Association Record*, 73(12), 228-230.
- Arunachalam, B., Henneberry, S. R., Lusk, J. L., & Norwood, F. B. (2009). An empirical investigation into the excessive-choice effect. *American journal of agricultural economics*, 91(3), 810-825.



## REFERENCES

---

- Aula, A., Jhaveri, N., & Käki, M. (2005). Information search and re-access strategies of experienced web users. *In Proceedings of the 14th international conference on World Wide Web* (pp. 583-592). ACM.
- Bar-Ilan, J., Keenoy, K., Yaari, E., & Levene, M. (2007). User rankings of search engine results. *Journal of the American Society for Information Science and Technology*, 58(9), 1254-1266.
- Barnette, J. J. (2000). Effects of stem and Likert response option reversals on survey internal consistency: If you feel the need, there is a better alternative to using those negatively worded stems. *Educational and Psychological Measurement*, 60(3), 361-370.
- Bates, M. J. (1979). Idea tactics. *Journal of the American Society for Information Science*, 30(5), 280-289.
- Bechwati, N. N., & Xia, L. (2003). Do computers sweat? The impact of perceived effort of online decision aids on consumers' satisfaction with the decision process. *Journal of Consumer Psychology*, 13(1), 139-148.
- Belkin, N. J., Cool, C., Stein, A., & Thiel, U. (1995). Cases, scripts, and information-seeking strategies: On the design of interactive information retrieval systems. *Expert systems with applications*, 9(3), 379-395.
- Benartzi, S., & Thaler, R. H. (2001). Naive diversification strategies in defined contribution saving plans. *American economic review*, 79-98.
- Bernard, M., Baker, R., & Fernandez, M. (2002). Paging vs. scrolling: Looking for the best way to present search results. *Usability News*, 4(1).
- Bharati, P., & Chaudhury, A. (2004). An empirical investigation of decision-making satisfaction in web-based decision support systems. *Decision support systems*, 37(2), 187-197.
- Bilal, D. (2000). Children's use of the Yahoo!igans! Web search engine: I. Cognitive, physical, and affective behaviors on fact-based search tasks. *Journal of the American Society for information Science*, 51(7), 646-665.

- 
- Björneborn, L. (2010). Design dimensions enabling divergent behaviour across physical, digital, and social library interfaces. *In Persuasive Technology* (pp. 143-149). Springer Berlin Heidelberg.
- Blandford, A., & Attfield, S. (2010). Interacting with information. *Synthesis Lectures on Human-Centered Informatics*, 3(1), 1-99.
- Botti, S. (2004). The psychological pleasure and pain of choosing: when people prefer choosing at the cost of subsequent outcome satisfaction. *Journal of personality and social psychology*, 87(3), 312-326.
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative research in psychology*, 3(2), 77-101.
- Brin, S., & Page, L. (1998). The anatomy of a large-scale hypertextual Web search engine. *Computer networks and ISDN systems*, 30(1), 107-117.
- Broder, A. (2002). A taxonomy of web search. *In ACM Sigir forum*. 36, pp. 3-10. ACM.
- Broniarczyk, S. M., Hoyer, W. D., & McAlister, L. (1998). Consumers' perceptions of the assortment offered in a grocery category: The impact of item reduction. *Journal of Marketing Research*, 166-176.
- Brumby, D. P., & Howes, A. (2003). Interdependence and past experience in menu choice assessment. *In the Proceedings of the 25th Annual Conference of the Cognitive Science Society* (p. 1320). Boston: LAWRENCE ERLBAUM ASSOC PUBL.
- Brumby, D. P., & Howes, A. (2008). Strategies for guiding interactive search: An empirical investigation into the consequences of label relevance for assessment and selection. *Human-Computer Interaction*, 23(1), 1-46.
- Campos, J., & Figueiredo, A. D. (2001). Searching the unsearchable: Inducing serendipitous insights. *In Case-based reasoning: Workshop program at ICCBR-2001*.
- Capra, R. (2010). HCI browser: A tool for studying web search behavior. *Proceedings of the American Society for Information Science and Technology*, 47(1), 1-2.

## REFERENCES

---

- Cardozo, R. N. (1965). An experimental study of customer effort, expectation, and satisfaction. *Journal of marketing research*, 244-249.
- Carroll, L. S., White, M. P., & Pahl, S. (2011). The impact of excess choice on deferment of decisions to volunteer. *Judgment and Decision Making*, 6(7), 629-637.
- Chernev, A. (2006). Decision focus and consumer choice among assortments. *Journal of Consumer Research*, 33(1), 50-59.
- Cho, J., Roy, S., & Adams, R. E. (2005). Page quality: In search of an unbiased web ranking. In *Proceedings of the 2005 ACM SIGMOD international conference on Management of data* (pp. 551-562). ACM.
- Churchill, G. A., & Surprenant, C. (1982). An investigation into the determinants of customer satisfaction. *Journal of marketing research*, 491-504.
- Cole, M., Gwizdka, J., Belkin, N., & Liu, C. (2011). User domain knowledge and eye movement patterns during search. In *Proceedings of the fifth workshop on human-computer interaction and information retrieval*.
- Collinsdictionary.com*. (2013). Retrieved August 1, 2013, from Collins English Thesaurus: <http://www.collinsdictionary.com/english-thesaurus>
- Cooksey, E. B. (2004). Too Important to Be Left to Chance—Serendipity and the Digital Library. *Science & Technology Libraries*, 25(1-2), 23-32.
- Corbin, J., & Strauss, A. (2008). *Basics of qualitative research: Techniques and procedures for developing grounded theory*. Sage.
- Cox, A. L., & Young, R. M. (2004). A rational model of the effect of information scent on the exploration of menus. In M. C. Lovett, C. Lebiere, & P. Munro, *Proceedings of the sixth international conference on cognitive modeling* (pp. 82-87). Lawrence Erlbaum Assoc.
- Crescenzi, A., Capra, R., & Arguello, J. (2013). Time pressure, user satisfaction and task difficulty. *The American Society for Information Science and Technology*, 50(1), 1-4.

- 
- Cummins, R. A., & Gullone, E. (2000). Why we should not use 5-point Likert scales: The case for subjective quality of life measurement. *In Proceedings of second international conference on quality of life in cities*, (pp. 74-93).
- Dawes, J. (2008). Do data characteristics change according to the number of scale points used? An experiment using 5 point, 7 point and 10 point scales. *Journal of Market Research*, 51(1).
- Dhar, R., & Nowlis, S. M. (1999). The effect of time pressure on consumer choice deferral. *Journal of Consumer Research*, 25(4), 369-384.
- Dictionary, O. E. (2013). Retrieved September 1, 2013, from Oxford English dictionary online: <http://www.oxforddictionaries.com/definition/english/serendipity>
- Diehl, K. (2005). When two rights make a wrong: Searching too much in ordered environments. *Journal of Marketing Research*, 42(3), 313-322.
- Dou, Z., Hu, S., Chen, K., Song, R., & Wen, J. R. (2011). Multi-dimensional search result diversification. *In Proceedings of the fourth ACM international conference on Web search and data mining* (pp. 475-484). ACM.
- Drosou, M., & Pitoura, E. (2010). Search result diversification. *ACM SIGMOD Record*, 39(1), 41-47.
- Eisenberg, M., & Barry, C. (1988). Order effects: A study of the possible influence of presentation order on user judgments of document relevance. *Journal of the American Society for Information Science*, 39(5), 293-300.
- Ellis, D. (1993). Modeling the information-seeking patterns of academic researchers: A grounded theory approach. *The Library Quarterly*, 63(4), 469-486.
- Erdelez, S. (1997). Information encountering: a conceptual framework for accidental information discovery. *In Proceedings of an international conference on Information seeking in context* (pp. 412-421). Taylor Graham Publishing.
- Fasolo, B., McClelland, G. H., & Todd, P. M. (2007). Escaping the tyranny of choice: When fewer attributes make choice easier. *Marketing Theory*, 7, 13-26.

## REFERENCES

---

- Fields, B., Keith, S., & Blandford, A. (2005). Designing for expert information finding strategies. In *People and computers XVIII—Design for life* (pp. 89-102). London: Springer .
- Foster, A., & Ford, N. (2003). Serendipity and information seeking: an empirical study. *Journal of Documentation*, 59(3), 321-340.
- Fox, C. R., Ratner, R. K., & Lieb, D. S. (2005). How subjective grouping of options influences choice and allocation: diversification bias and the phenomenon of partition dependence. *Journal of Experimental Psychology: General*, 134(4), 538.
- Fu, W. T., & Pirolli, P. (2007). SNIF-ACT: A cognitive model of user navigation on the World Wide Web. *Human-Computer Interaction*, 22(4), 355-412.
- Gao, J., Yuan, W., Li, X., Deng, K., & Nie, J. Y. (2009). Smoothing clickthrough data for web search ranking. In *Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval* (pp. 355-362). ACM.
- Gatian, A. W. (1994). Is user satisfaction a valid measure of system effectiveness? *Information & Management*, 26(3), 119-131.
- Geers, A. L., Rose, J. P., Rasinski, H. M., Brown, J. A., & Helfer, S. G. (2013). Why does choice enhance treatment effectiveness? Using placebo treatments to demonstrate the role of personal control. *Journal of Personality and Social Psychology*, 105(4), 549-566.
- Giese, J. L., & Cote, J. A. (2000). Defining consumer satisfaction. *Academy of marketing science review*, 1(1), 1-22.
- Goldman, E. (2006). Search engine bias and the demise of search engine utopianism. *Yale Journal of Law & Technology*, 6-8.
- Gordon, M., & Pathak, P. (1999). Finding information on the World Wide Web: the retrieval effectiveness of search engines. *Information Processing & Management*, 35(2), 141-180.

- 
- Guan, Z., & Cutrell, E. (2007). An eye tracking study of the effect of target rank on web search. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 417-420). ACM.
- Hahn, M., Lawson, R., & Lee, Y. G. (1992). The effects of time pressure and information load on decision quality. *Psychology & Marketing*, 9(5), 365-378.
- Hargittai, E., & Hinnant, A. (2008). Digital inequality differences in young adults' use of the internet. *Communication Research*, 35(5), 602-621.
- Hauff, C., & Houben, G. J. (2012). Serendipitous browsing: Stumbling through wikipedia. In *Searching4Fun! workshop (co-located with ECIR 2012)*.
- Haynes, G. A. (2009). Testing the boundaries of the choice overload phenomenon: The effect of number of options and time pressure on decision difficulty and satisfaction. *Psychology & Marketing*, 26(3), 204-212.
- Heitmann, M., Lehmann, D. R., & Herrmann, A. (2007). Choice goal attainment and decision and consumption satisfaction. *Journal of Marketing Research*, 44(2), 234-250.
- Herner, S. (1970). Browsing. In A. Kent, & H. Lancour, *Encyclopaedia of Library and Information Science* (Vol. 3, pp. 408-415). New York: Dekker.
- Hölscher, C., & Strube, G. (2000). Web search behavior of Internet experts and newbies. *Computer networks*, 33(1), 337-346.
- Hsieh-Yee, I. (1993). Effects of search experience and subject knowledge on the search tactics of novice and experienced searchers. *Journal of the American Society for Information Science*, 44(3), 161-174.
- Huang, M. H., & Wang, H. Y. (2004). The influence of document presentation order and number of documents judged on users' judgments of relevance. *Journal of the American Society for Information Science and Technology*, 55(11), 970-979.
- Huffman, C., & Kahn, B. E. (1998). Variety for sale: mass customization or mass confusion? *Journal of retailing*, 74(4), 491-513.

## REFERENCES

---

- Iaquinta, L., De Gemmis, M., Lops, P., Semeraro, G., Filannino, M., & Molino, P. (2008). Introducing serendipity in a content-based recommender system. *In Eighth International Conference on Hybrid Intelligent Systems, HIS'08*. (pp. 168-173). IEEE.
- Iyengar, S. S., & Lepper, M. R. (2000). When choice is demotivating: Can one desire too much of a good thing? *Journal of personality and social psychology*, 79(6), 995-1006.
- Iyengar, S. S., Wells, R. E., & Schwartz, B. (2006). Doing Better but Feeling Worse Looking for the “Best” Job Undermines Satisfaction. *Psychological Science*, 17(2), 143-150.
- Jansen, B. J., & Spink, A. (2006). How are we searching the World Wide Web? A comparison of nine search engine transaction logs. *Information Processing & Management*, 42(1), 248-263.
- Jansen, B. J., Spink, A., & Saracevic, T. (2000). Real life, real users, and real needs: a study and analysis of user queries on the web. *Information processing & management*, 36(2), 207-227.
- Kahn, B. E. (1995). Consumer variety-seeking among goods and services: An integrative review. *Journal of Retailing and Consumer Services*, 2(3), 139-148.
- Kahn, B. E., & Isen, A. M. (1993). The influence of positive affect on variety seeking among safe, enjoyable products. *Journal of Consumer Research*, 257-270.
- Kahn, B. E., & Wansink, B. (2004). The influence of assortment structure on perceived variety and consumption quantities. *Journal of Consumer Research*, 30(4), 519-533.
- Kahn, B. E., Weingarten, E., & Townsend, C. (2013). Assortment Variety: Too Much of a Good Thing? *Review of Marketing Research*, 10, 1-23.
- Keane, M. T., O'Brien, M., & Smyth, B. (2008). Are people biased in their use of search engines? *Communications of the ACM*, 51(2), 49-52.

- 
- Kim, K. S., & Allen, B. (2002). Cognitive and task influences on Web searching behavior. *Journal of the American Society for Information Science and Technology*, 53(2), 109-119.
- Kristensen, J. (1993). Expanding end-users' query statements for free text searching with a search-aid thesaurus. *Information Processing & Management*, 29(6), 733-744.
- Likert, R. (1932). A technique for the measurement of attitudes. *Archives of psychology*.
- Liu, C., Gwizdka, J., Liu, J., Xu, T., & Belkin, N. J. (2010). Analysis and evaluation of query reformulations in different task types. *Proceedings of the American Society for Information Science and Technology*, 47(1), 1-9.
- Loewenstein, G. (1999). Is more choice always better? *Social Security Brief*, 7, 1-8.
- Lorigo, L., Haridasan, M., Brynjarsdóttir, H., Xia, L., Joachims, T., Gay, G., et al. (2008). Eye tracking and online search: Lessons learned and challenges ahead. *Journal of the American Society for Information Science and Technology*, 59(7), 1041-1052.
- Makri, S., & Blandford, A. (2012a). Coming across information serendipitously – Part 1: A process model. *Journal of Documentation*, 68(5), 684 – 705.
- Makri, S., & Blandford, A. (2012b). Coming across information serendipitously – Part 2: A classification framework. *Journal of Documentation*, 68(5), 706 – 724.
- Makri, S., Blandford, A., Woods, M., Sharples, S., & Maxwell, D. (2014). “Making my own luck”: Serendipity strategies and how to support them in digital information environments. *Journal of the Association for Information Science and Technology*.
- Malhotra, N. K. (1982). Information load and consumer decision making. *Journal of consumer research*, 419-430.
- Marchionini, G. (1997). *Information seeking in electronic environments*. Cambridge University Press.



## REFERENCES

---

- Martin, K., & Quan-Haase, A. (2014). Changing Chance Encounters: Historians, Serendipity, and the Digital Text. *In Proceedings of the Annual Conference of the Canadian Association for Information Science*.
- McAlister, L., & Pessemier, E. (1982). Variety seeking behavior: An interdisciplinary review. *Journal of Consumer research*, 311-322.
- McCay-Peet, L., & Toms, E. (2011). Measuring the dimensions of serendipity in digital environments. *Information Research: An International Electronic Journal*, 16(3).
- McCay-Peet, L., & Toms, E. G. (2010). The process of serendipity in knowledge work. *In Proceedings of the third symposium on Information interaction in context* (pp. 377-382). ACM.
- Menon, S., & Kahn, B. E. (1995). The impact of context on variety seeking in product choices. *Journal of Consumer Research*, 285-295.
- Messner, C., & Wänke, M. (2011). Unconscious information processing reduces information overload and increases product satisfaction. *Journal of Consumer Psychology*, 21(1), 9-13.
- Miller, G. A. (1995). WordNet: a lexical database for English. *Communications of the ACM*, 38(11), 39-41.
- Mogilner, C., Rudnick, T., & Iyengar, S. S. (2008). The mere categorization effect: How the presence of categories increases choosers' perceptions of assortment variety and outcome satisfaction. *Journal of Consumer Research*, 35(2), 202-215.
- Narayana, C. L., & Markin, R. J. (1975). Consumer behavior and product performance: An alternative conceptualization. *The Journal of Marketing*, 1-6.
- Nutefall, J. E., & Ryder, P. M. (2010). The serendipitous research process. *The Journal of Academic Librarianship*, 36(3), 228-234.
- Oliver, R. L. (1996). Varieties of value in the consumption satisfaction response. *Advances in consumer research*, 23, 143-147.
- Oliver, R. L., & DeSarbo, W. S. (1988). Response determinants in satisfaction judgments. *Journal of consumer research*, 495-507.

- 
- Oliver, R. L., Rust, R. T., & Varki, S. (1997). Customer delight: foundations, findings, and managerial insight. *Journal of Retailing*, 73(3), 311-336.
- Oppewal, H., & Koelemeijer, K. (2005). More choice is better: Effects of assortment size and composition on assortment evaluation. *International Journal of Research in Marketing*, 22(1), 45-60.
- Oulasvirta, A., Hukkinen, J. P., & Schwartz, B. (2009). When more is less: the paradox of choice in search engine use. In *proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval*, (pp. 516-523).
- Oxforddictionaries.com* . (2013). Retrieved August 1, 2013, from Oxford Dictionaries: <http://www.oxforddictionaries.com/thesaurus/>
- Page, L., Brin, S., Motwani, R., & Winograd, T. (1999). *The PageRank citation ranking: Bringing order to the web*. Technical Report, Stanford InfoLab.
- Pan, B., Hembrooke, H., Joachims, T., Lorigo, L., Gay, G., & Granka, L. (2007). In Google we trust: Users' decisions on rank, position, and relevance. *Journal of Computer-Mediated Communication*, 12(3), 801-823.
- Pandey, S., Roy, S., Olston, C., Cho, J., & Chakrabarti, S. (2005). Shuffling a stacked deck: the case for partially randomized ranking of search engine results. In *Proceedings of the 31st international conference on Very large data bases*, (pp. 781-792).
- Park, J. Y., & Jang, S. S. (2013). Confused by too many choices? Choice overload in tourism. *Tourism Management*, 35, 1-12.
- Patall, E. A. (2013). Constructing motivation through choice, interest, and interestingness. *Journal of Educational Psychology*, 105(2), 522-534.
- Patall, E. A., Cooper, H., & Wynn, S. R. (2010). The effectiveness and relative importance of choice in the classroom. *Journal of Educational Psychology*, 102(4), 896-915.

## REFERENCES

---

- Piolat, A., Roussey, J. Y., & Thunin, O. (1997). Effects of screen presentation on text reading and revising. *International Journal of Human-Computer Studies*, 47(4), 565-589.
- Purgailis Parker, L. M., & Johnson, R. E. (1990). Does order of presentation affect users' judgment of documents? *Journal of the American Society for Information Science*, 41(7), 493-494.
- Raghavan, V., Bollmann, P., & Jung, G. S. (1989). A critical investigation of recall and precision as measures of retrieval system performance. *ACM Transactions on Information Systems (TOIS)*, 7(3), 205-229.
- Raju, P. S. (1980). Optimum stimulation level: its relationship to personality, demographics, and exploratory behavior. *Journal of consumer research*, 272-282.
- Ratner, R. K., Kahn, B. E., & Kahneman, D. (1999). Choosing less-preferred experiences for the sake of variety. *Journal of Consumer Research*, 26(1), 1-15.
- Reibstein, D. J., Youngblood, S. A., & Fromkin, H. L. (1975). Number of choices and perceived decision freedom as a determinant of satisfaction and consumer behavior. *Journal of Applied Psychology*, 60(4), 434-437.
- Reneker, M. H. (1993). A qualitative study of information seeking among members of an academic community: methodological issues and problems. *The Library Quarterly*, 63(4), 487-507.
- Reutskaja, E., & Hogarth, R. M. (2009). Satisfaction in choice as a function of the number of alternatives: When “goods satiate”. *Psychology & Marketing*, 26(3), 197-203.
- Rice, R. E., McCreddie, M. M., & Chang, S. L. (2001). *Accessing and browsing information and communication*. MIT Press.
- Rieger, O. Y. (2009). Search engine use behavior of students and faculty: User perceptions and implications for future research. *First Monday*, 14(12).

- 
- Rieh, S. Y. (2004). On the Web at home: Information seeking and Web searching in the home environment. *Journal of the American Society for Information Science and Technology*, 55(8), 743-753.
- Roberts, R. M. (1989). *Serendipity: Accidental discoveries in science*. New York: Wiley.
- Rolls, B. J., Rowe, E. A., Rolls, E. T., Kindston, B., Megson, A., & Gunary, R. (1981). Variety in a Meal Enhances Food Intake in Man. *Physiology and Behavior*, 26, 215–221.
- Rose, D. E., & Levinson, D. (2004). Understanding user goals in web search. In *Proceedings of the 13th international conference on World Wide Web* (pp. 13-19). ACM.
- Rosenman, M. (1988). Serendipity and scientific discovery. *The Journal of Creative Behavior*, 22(2), 132-138.
- Sainfort, F., & Booske, B. C. (2000). Measuring post-decision satisfaction. *Medical Decision Making*, 20(1), 51-61.
- Santos, R. L., Macdonald, C., & Ounis, I. (2010). Exploiting query reformulations for web search result diversification. In *Proceedings of the 19th international conference on World wide web* (pp. 881-890). ACM.
- Sauro, J., & Lewis, J. R. (2011). When designing usability questionnaires, does it hurt to be positive? In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 2215-2224). ACM.
- Scheibehenne, B., Greifeneder, R., & Todd, P. M. (2009). What moderates the too-much-choice effect? *Psychology & Marketing*, 26(3), 229-253.
- Scheibehenne, B., Greifeneder, R., & Todd, P. M. (2010). Can There Ever Be Too Many Options? A Meta-Analytic Review of Choice Overload. *Journal of Consumer Research*, 37(3), 409-425.
- Schwartz, B. (2004). *The paradox of choice: Why less is more*. New York: Ecco.

## REFERENCES

---

- Schwartz, B., Ward, A., Monterosso, J., Lyubomirsky, S., White, K., & Lehman, D. R. (2002). Maximizing versus satisficing: happiness is a matter of choice. *Journal of personality and social psychology*, 83(5), 1178-1197.
- Schwarz, E., Beldie, I. P., & Pastoor, S. (1983). A Comparison of Paging and Scrolling for Changing Screen Contents by Inexperienced Users. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 25(3), pp. 279-282.
- Shocker, A. D., Ben-Akiva, M., Boccara, B., & Nedungadi, P. (1991). Consideration set influences on consumer decision-making and choice: Issues, models, and suggestions. *Marketing letters*, 2(3), 181-197.
- Simon, H. A. (1955). A behavioral model of rational choice. *Quarterly Journal of Economics*, 69, 99-118.
- Simonson, I. (1990). The effect of purchase quantity and timing on variety-seeking behavior. *Journal of Marketing Research*, 150-162.
- Simonson, I., & Winer, R. S. (1992). The influence of purchase quantity and display format on consumer preference for variety. *Journal of Consumer Research*, 133-138.
- Söderlund, M. (2002). Customer familiarity and its effects on satisfaction and behavioral intentions. *Psychology & Marketing*, 19(10), 861-879.
- Spink, A., Wolfram, D., Jansen, M. B., & Saracevic, T. (2001). Searching the web: The public and their queries. *Journal of the American society for information science and technology*, 52(3), 226-234.
- Stelmaszewska, H., & Blandford, A. (2002). Patterns of interactions: user behaviour in response to search results. In A. Blandford, & G. Buchanan, *Proceedings of the JCDL Workshop on Usability of Digital Libraries Usability of Digital Libraries 2002* (pp. 29 - 32).
- Thatcher, A. (2008). Web search strategies: The influence of Web experience and task type. *Information Processing & Management*, 44(3), 1308-1329.

- 
- Thesaurus.com*. (2013). Retrieved August 1, 2013, from Roget's 21st Century Thesaurus, Third Edition: <http://thesaurus.com/>
- Tintarev, N., Dennis, M., & Masthoff, J. (2013). Adapting Recommendation Diversity to Openness to Experience: A Study of Human Behaviour. In *User Modeling, Adaptation, and Personalization* (pp. 190-202). Springer Berlin Heidelberg.
- Toms, E. G. (2000). Serendipitous Information Retrieval. In *DELOS Workshop: Information Seeking, Searching and Querying in Digital Libraries*.
- Toms, E. G., & McCay-Peet, L. (2009). Chance encounters in the digital library. In *Research and Advanced Technology for Digital Libraries*. Springer Berlin Heidelberg.
- Tse, D. K., & Wilton, P. C. (1988). Models of consumer satisfaction formation: an extension. *Journal of marketing research*, 204-212.
- Van Andel, P. (1994). Anatomy of the unsought finding. serendipity: Origin, history, domains, traditions, appearances, patterns and programmability. *The British Journal for the Philosophy of Science*, 45(2), 631-648.
- Van Trijp, H. C., Hoyer, W. D., & Inman, J. J. (1996). Why switch? Product category: level explanations for true variety-seeking behavior. *Journal of Marketing Research*, 281-292.
- Vandenbosch, B., & Huff, S. L. (1997). Searching and scanning: how executives obtain information from executive information systems. *MIS Quarterly*, 21(1), 81-107.
- Wang, Q., & Shukla, P. (2013). Linking sources of consumer confusion to decision satisfaction: the role of choice goals. *Psychology & Marketing*, 30(4), 295-304.
- Ward, G. (1996). Moby thesaurus.
- Westbrook, R. A., & Oliver, R. L. (1991). The dimensionality of consumption emotion patterns and consumer satisfaction. *Journal of consumer research*, 84-91.
- White, R. W., & Morris, D. (2007). Investigating the querying and browsing behavior of advanced search engine users. In *Proceedings of the 30th annual international*

## REFERENCES

---

- ACM SIGIR conference on Research and development in information retrieval* (pp. 255-262). ACM.
- White, R. W., Dumais, S. T., & Teevan, J. (2009). Characterizing the influence of domain expertise on web search behavior. *In Proceedings of the Second ACM International Conference on Web Search and Data Mining* (pp. 132-141). ACM.
- Xie, I., & Joo, S. (2010). Transitions in search tactics during the Web-based search process. *Journal of the American Society for Information Science and Technology*, 61(11), 2188-2205.
- Yu, C., Lakshmanan, L., & Amer-Yahia, S. (2009). It takes variety to make a world: diversification in recommender systems. *In Proceedings of the 12th international conference on extending database technology: Advances in database technology* (pp. 368-378). ACM.
- Zhang, S., & Fitzsimons, G. J. (1999). Choice-process satisfaction: The influence of attribute alignability and option limitation. *Organizational Behavior and Human Decision Processes*, 77(3), 192-214.
- Zheng, W., Wang, X., Fang, H., & Cheng, H. (2012). Coverage-based search result diversification. *In Proceedings of the 19th international conference on World wide web* (pp. 881-890). ACM.
- Zucker, D., & Remaud, H. (2014). Does Choice Overload Exist in Wine Retail? *In Proceedings of 8th Academy of Wine Business Research Conference*.

## APPENDIX A

### ETHICS CHECKLIST

This ethics checklist was reviewed prior to each experiment. The content of the checklist is based on Department of Computer Science 13-point ethics checklist. Our responses reports below:

#### UNIVERSITY OF BATH, Department of Computer Science

#### 13-POINT ETHICS CHECK LIST

This document describes the 13 issues that need to be considered carefully before students or staffs involve other people (“participants”) for the collection of information as part of their project or research.

1. *Have you prepared a briefing script for volunteers?*

*You must explain to people what they will be required to do, the kind of data you will be collecting from them and how it will be used.*

- Yes, a briefing script were prepared and included in a consent form in all experiments and a study. All participants are aware that they can withdraw from the experiment at any time without any consequence.

2. *Will the participants be using any non-standard hardware?*

*Participants should not be exposed to any risks associated with the use of non-standard equipment: anything other than pen and paper or typical interaction with PCs on desks is considered non-standard.*

- No, participants will be using a standard hardware to perform interaction in a laboratory. Both hardware and software used in all experiments pose no



immediate physical or mental risks to participants.

3. *Is there any intentional deception of the participants?*

*Withholding information or misleading participants is unacceptable if participants are likely to object or show unease when debriefed.*

- In Experiment 2, participants were informed that one of three processes was used to rank the returns of a search. After completing the experiment, participants were debriefed during the post-experiment session that in all cases the Google ranking was preserved. None of participants in Experiment 2 were likely to object or show unease when debriefed.
- All other experiments and the study did not use deception but the purpose of the investigation was withheld until participants completed the experimental tasks. This is essential to prevent the impact of our main objective on participants' behavior. All participants were given full debriefing about the main purpose of the study after they had completed the experiments.

4. *How will participants voluntarily give consent?*

*If the results of the evaluation are likely to be used beyond the term of the project (for example, the software is to be deployed, or the data is to be published), then signed consent is necessary. A separate consent form should be signed by each participant.*

- All experiments used paper-based consent forms. The consent forms included the detail of the experiment, how their data will be used, and their right to withdraw from the experiment.

5. *Will the participants be exposed to any risks greater than those encountered in their normal work life?*

*Investigators have a responsibility to protect participants from physical and mental harm during the investigation. The risk of harm must be no greater than in ordinary life.*

- No, participants will not be exposed to any risks greater than those encountered in their normal work life.

6. *Are you offering any incentive to the participants?*

*The payment of participants must not be used to induce them to risk harm beyond that which they risk without payment in their normal lifestyle.*

- 
- Yes, there will be monetary incentives for participating. However the incentive offered do not encourage participants to risk physical or mental harm beyond that which they risk in everyday life. Such incentives were used for recruiting purpose, yet no participants were forced to take part in our experiments.
7. *Are any of your participants under the age of 16?*  
*Parental consent is required for participants under the age of 16.*
- No.
8. *Do any of your participants have an impairment that will limit their understanding or communication?*  
*Additional consent is required for participants with impairments.*
- No.
9. *Are you in a position of authority or influence over any of your participants?*  
*A position of authority or influence over any participant must not be allowed to pressurise participants to take part in, or remain in, any experiment.*
- No.
10. *Will the participants be informed that they could withdraw at any time?*  
*All participants have the right to withdraw at any time during the investigation.*  
*They should be told this in the introductory script.*
- Yes, participants will be informed about their right to withdraw at any time.  
This is written explicitly clear in the consent forms.
11. *Will the participants be informed of your contact details?*  
*All participants must be able to contact the investigator after the investigation. They should be given the details of the Unit Lecturer or Supervisor as part of the debriefing.*
- Yes, the contact detail is given in the consent forms.
12. *Will participants be de-briefed?*  
*The student must provide the participants with sufficient information in the debriefing to enable them to understand the nature of the investigation.*
- Yes, all participants will be debriefed after they completed the experiment.
13. *Will the data collected from the participants be stored in an anonymous form?*

## APPENDIX A ETHICS CHECKLIST

---

*All participant data (hard copy and soft copy) should be stored securely, and in anonymous form.*

- The data collected from participants (text files and computer screen recordings) will be stored in anonymous form. This data will be kept in a secure location and will not be distributed or disclosed to the members of the public.

## **APPENDIX B**

### **EXPERIMENT 1 MATERIALS**

#### **B.1 CONSENT FORM**

##### **STUDY OVERVIEW**

This is a study of search engine usage. The task you will be working on requires you to find information for 10 search tasks using Google. You will be asked questions after each task as well as at the end of each session.

The quality of the websites you find will be evaluated in order to allocate the prizes. There will be three prizes of £20 in cash for the best overall performance.

##### **IMPORTANT INFORMATION**

- Your results will not be passed to any third party and are not being collected for commercial reasons.
- Participation in this study does not involve physical or mental risks outside of those encountered in everyday life.
- All procedures and information can be taken at face value and no deception is involved.
- You have the right to withdraw from the study at any time and to have any data about you destroyed. If you do decide to withdraw, please inform the experimenter.
- You will be reimbursed for your time upon successfully completing the study.

By signing this form you acknowledge that you have read the information given above and understand the terms and conditions of this study.

Age :	Gender :
Occupation :	Department :
Signed :	Date :

Experimenter: Pawitra Chiravirakul, Dept. of Computer Science.

PC345@BATH.AC.UK

## B.2 TASK MATERIALS

### INSTRUCTION

In this study, you will be required to find an article that relates to your research topic by using Google. During the search task you will be provided with two lists of search results; Keyword Search list and Synonym Search list.

This program will guide you through a training session before performing your search task.

In order to perform the task, you can search and browse web pages with no time limit. You will complete the task by submitting your selected article.

You will be required to rate your selected article and answer the questionnaire at the end.

### TASK QUESTION SETS

#### *Training Question*

- Why did the ancient Egyptians mummify their dead?

#### *Task Question Set A*

1. What determines the cost of living in UK?
2. Why did The King's Speech win the best director Oscar for Tom Hooper?
3. What is the US President's opinion about the killing of Bin Laden by US forces?
4. How did Facebook become the most popular social networking website?
5. What caused the death of whales on an Irish beach this year?
6. Why is meditation sometimes recommended for managing stress?

- 
7. How could we measure "Happiness"?
  8. How does association football's current popularity relate to its historic popularity?
  9. What caused the current conflict between North and South Korea?
  10. What determines the value of a house in the UK?

*Task Question Set B*

1. What should be the criteria of selecting a healthy diet?
2. Why is Internet shopping likely to grow substantially in the recent future?
3. What do critics most admire about the movie Avatar?
4. How was the first elevator invented?
5. What determines currency exchange rates?
6. Why is China famous for pottery?
7. Why are alcoholic drinks sold only to people age 18 or over?
8. Why should we recycle?
9. Why at present is the divorce rate increasing in all European countries?
10. Why have UK University tuition fees been raised?



## **APPENDIX C**

### **EXPERIMENT 2 MATERIALS**

#### **C.1 CONSENT FORM**

##### **STUDY OVERVIEW**

This is a study of search engine usage. The task you will be working on requires you to select web pages from lists of search results, in order to answer 12 separate search questions. The quality of the websites you find will be evaluated in order to allocate the prizes. There will be three prizes of £30 in cash for the best overall performance. For the participation today you will be paid £5 cash at the end of the study.

##### **IMPORTANT INFORMATION**

- Your recorded results will not be passed to any third party and are not being collected for commercial reasons.
- Participation in this study does not involve physical or mental risks outside of those encountered in everyday life.
- All procedures and information can be taken at face value and no deception is involved.
- You have the right to withdraw from the study at any time and to have any data about you destroyed. If you do decide to withdraw, please inform the experimenter.
- You will be reimbursed for your time upon successfully completing the study.



By signing this form you acknowledge that you have read the information given above and understand the terms and conditions of this study.

Age :	Gender :
Occupation :	Department :
Signed :	Date :

Experimenter: Pawitra Chiravirakul, Dept. of Computer Science.

PC345@BATH.AC.UK

### C.2 TASK MATERIALS

#### INSTRUCTION

In this study, you will be required to select web pages from lists of search results, in order to answer 12 separate search questions.

For each task you will see the Question you want to answer and the search term that was used.

The 12 tasks are divided into 3 sessions. In each session, the lists of search results will be ranked in one of three ways: Expert, Novice, or Random. The way these rankings work will be explained later.

This program will guide you through the tasks. For each task:

- You must use the snippets to choose a single link. You can then inspect the web page that is linked to.
- You must submit your single chosen link as an answer.
- You will then be asked to rate your selected web page

At end of each session you will be asked to make some further ratings.

#### TASK QUESTION SETS

##### *Training Question*

- Why did the ancient Egyptians mummify their dead?  
Search Term: Egyptians mummification

---

*Task Question Set 1*

1. What determines the cost of living in the UK?  
Search Term: determines cost of living in UK
2. How do penguins survive in their habitat?  
Search Term: penguins survival method
3. Why is meditation sometimes recommended for managing stress?  
Search Term: meditation for managing stress
4. How was the first elevator invented?  
Search Term: first elevator invented

*Task Question Set 2*

1. What determines the value of a house in the UK?  
Search Term: value of house in UK measurement
2. What do critics most admire about the movie Avatar?  
Search Term: critics most admired Avatar 2009
3. What should be the criteria of selecting a healthy diet?  
Search Term: criteria of selecting healthy diet
4. How does association football UK is current popularity relate to its historic popularity?  
Search Term: popularity of association football UK current and historic

*Task Question Set 3*

1. What determines currency exchange rates?  
Search Term: currency exchange rate factor
2. How did Facebook become the popular social networking website?  
Search Term: Facebook popularity
3. Why should we recycle?  
Search Term: why recycle
4. Why China is famous for pottery?  
Search Term: why Chinese pottery famous



## **APPENDIX D**

### **EXPERIMENT 3 MATERIALS**

#### **D.1 CONSENT FORM**

##### **STUDY OVERVIEW**

This is a study of search engine usage. The task you will be working on requires you to select web pages from lists of search results, in order to answer 16 separate search questions, which are divided into 4 sessions. The lists of search results will be ranked in one of the two ways, Google ranking or Random ranking. You will be asked questions after each task as well as the end of each session.

The quality of the websites you find will be evaluated in order to allocate the prizes. There will be three prizes of £30 in cash for the best overall performance.

For the participation today you will be paid £5 cash at the end of the study.

##### **IMPORTANT INFORMATION**

- Your recorded results will not be passed to any third party and are not being collected for commercial reasons.
- Participation in this study does not involve physical or mental risks outside of those encountered in everyday life.
- All procedures and information can be taken at face value and no deception is involved.

## APPENDIX D EXPERIMENT 3 MATERIALS

---

- You have the right to withdraw from the study at any time and to have any data about you destroyed. If you do decide to withdraw, please inform the experimenter.
- You will be reimbursed for your time upon successfully completing the study.

By signing this form you acknowledge that you have read the information given above and understand the terms and conditions of this study.

Age :	Gender :
Occupation :	Department :
Signed :	Date :

Experimenter: Pawitra Chiravirakul, Dept. of Computer Science.  
PC345@BATH.AC.UK

### D.2 TASK MATERIALS

#### INSTRUCTION

In this study, you will be required to select web pages from lists of search results, in order to answer 16 separate search questions.

For each task you will see the Question you want to answer, the search term that was used and time that you have left to finish the task.

The 16 tasks are divided into 4 sessions. In each session:

- A number of search results are presented, as found by Google.
- These search results will be ranked in one of two ways: Google ranking or Random ranking.
- This program will guide you through the tasks. For each task:
- You must use the snippets to choose a single link. You can then inspect the web page that is linked to.
- You must submit your single chosen link as an answer within time provided.
- You will then be asked to rate your selected web page

---

At end of each session you will be asked to make some further ratings.

## **TASK QUESTION SETS**

### *Training Questions*

- Why did the ancient Egyptians mummify their dead?  
Search Term: Egyptians mummification
- How does economics affect our daily life?  
Search Term: economics affect daily life

### *Task Question Set 1*

1. What determines the cost of living in the UK?  
Search Term: determines cost of living in UK
2. How do penguins survive in their habitat?  
Search Term: Penguins survival method
3. Why is meditation sometimes recommended for managing stress?  
Search Term: meditation for managing stress
4. How was the first elevator invented?  
Search Term: first elevator invented

### *Task Question Set 2*

1. What determines the value of a house in the UK?  
Search Term: value of house in UK measurement
2. What do critics most admire about the movie Avatar?  
Search Term: critics most admired Avatar 2009
3. What should be the criteria of selecting a healthy diet?  
Search Term: criteria of selecting healthy diet
4. How does association football UK is current popularity relate to its historic popularity?  
Search Term: popularity of association football UK current and historic

### *Task Question Set 3*

1. What determines currency exchange rates?  
Search Term: currency exchange rate factor

2. How did Facebook become the popular social networking website?

Search Term: Facebook popularity

3. Why should we recycle?

Search Term: why recycle

4. Why China is famous for pottery?

Search Term: why Chinese pottery famous

*Task Question Set 4*

1. Why at present is the divorce rate increasing in all European countries?

Search Term: divorce rate increasing in Europe

2. What caused the current conflict between North and South Korea?

Search Term: cause of conflict north and South Korea

3. How could we measure Happiness?

Search Term: measure happiness

4. Why is Internet shopping likely to grow substantially in the recent future?

Search Term: Internet shopping future growth

## **APPENDIX E**

### **EXPERIMENT 4 MATERIALS**

#### **E.1 CONSENT FORM**

##### **STUDY OVERVIEW**

This is a study of search engine use. The task you will be working on requires you to select web pages from lists of search results, in order to answer 12 separate search questions. The lists of search results will be ranked by Google. Tasks are divided into 3 sessions (4 questions per session). Each session has different types of supplementary web pages: no supplementary, supplementary from Synonym search, and supplementary from the third quartile of a search result list. You will be asked questions after each task as well as the end of each session.

The quality of the websites you find will be evaluated in order to allocate the prizes. There will be three prizes of £30 in cash for the best overall performance. For the participation today you will be paid £5 cash at the end of the study.

##### **IMPORTANT INFORMATION**

- Your recorded results will not be passed to any third party and are not being collected for commercial reasons.
- Participation in this study does not involve physical or mental risks outside of those encountered in everyday life.
- All procedures and information can be taken at face value and no deception is involved.



- You have the right to withdraw from the study at any time and to have any data about you destroyed. If you do decide to withdraw, please inform the experimenter.
- You will be reimbursed for your time upon successfully completing the study.

By signing this form you acknowledge that you have read the information given above and understand the terms and conditions of this study.

Age :	Gender :
Occupation :	Department :
Signed :	Date :

Experimenter: Pawitra Chiravirakul, Dept. of Computer Science.  
PC345@BATH.AC.UK

### E.2 TASK MATERIALS

#### INSTRUCTION

In this study, you are required to select web pages from lists of search results, in order to answer 12 separate search questions. For each task you will see the Question you want to answer, the search term that was used and time that you have left to finish the task.

The lists of search results are ranked by Google. Tasks are divided into 3 sessions (4 questions per session). Each session has different types of supplementary web pages: no supplementary, supplementary from Synonym search, and supplementary from the third quartile of a search result list. These supplementary web pages are provided in addition to the original result list.

- *Supplementary from Synonym search*: the first noun found in a search term is substituted by its synonym before searching. A different synonym is used for each supplementary. A supplementary page is selected from the top of a synonym-search result list.
- *Supplementary from further down the result list*: supplementary web pages are selected from the third quartile and the forth quartile of a result list.

---

This program will guide you through the tasks. For each task:

- You must use the snippets to choose a single web page. You can then inspect the content of the selected web page that is linked to.
- You must choose your web page within the time allowed.
- You will then be asked to rate your selected web page
- At end of each session you will be asked to make some further ratings.

**Questionnaire after each task [LikertType]**

*For the single choice condition,*

1. You are satisfied with your selected page with respect to the information it provides concerning your question. [1=Least satisfied, 10=Most satisfied]
2. Your selected page was satisfying in content, independently of your question. [1=Least satisfied, 10=Most satisfied]
3. The total list of 48 web pages returned by the search was satisfactory. [1=Least satisfied, 10=Most satisfied]
4. The search and selection process was satisfactory. [1=Least satisfied, 10=Most satisfied]
5. I am confident I chose the best page from the available set. [1=Least confident, 10=Most confident]

*For the multiple choices condition,*

1. You are satisfied with your three selected pages with respect to the information they provide concerning your question. [1=Least satisfied, 10=Most satisfied]
2. Your three selected pages were satisfying in content, independently of your question. [1=Least satisfied, 10=Most satisfied]
3. The total list of 48 web pages returned by the search was satisfactory. [1=Least satisfied, 10=Most satisfied]
4. The search and selection process was satisfactory. [1=Least satisfied, 10=Most satisfied]
5. I am confident that I chose the best three web pages from the available set. [1=Least confident, 10=Most confident]

### Questionnaire after finishing all tasks

LikertType:

1. Overall, how satisfied are you with the pages you chose?  
[1=Least satisfied, 10=Most satisfied]
2. Overall, how satisfied are you with the lists of pages that the search engine returned?  
[1=Least satisfied, 10=Most satisfied]

OpenAnswer:

3. What was your strategy of choosing the web pages?
4. To what extent were the supplementary web pages provided useful?

### TASK QUESTION SETS

The following task questions were used in Experiment 4 and Experiment 5.

#### *Training Questions*

- How does economics affect our daily life?  
Search Term: economics affect daily life
- How can we manage time effectively?  
Search Term: how to manage time effectively

#### *Task Question Set*

1. What were the most important wars fought in the history of the UK?  
Search Term: The most important wars in the UK
2. What determines the price of real estate?  
Search Term: what determine prices of real estates
3. How can we measure "stress"?  
Search Term: how to measure stress
4. What were the major effects of World War II on European countries?  
Search Term: major effects of world war 2 in Europe
5. How was the first personal computer invented?  
Search Term: first personal computer invented
6. How can we learn a language successfully?  
Search Term: how to learn language successfully

- 
7. What is the likely future development of mobile phones?  
Search Term: future development of mobile phones
  8. What are the most effective ways of reducing the greenhouse effect?  
Search Term: most effective ways of reducing greenhouse effect
  9. How can we improve our own critical thinking skills?  
Search Term: how to improve our critical thinking skills
  10. What caused the recent economic crisis in Europe?  
Search Term: causes of recent economic crisis in Europe
  11. What is the most beautiful place to visit in Scotland?  
Search Term: most beautiful place to visit in Scotland
  12. How can we protect our privacy online?  
Search Term: how to protect privacy online



## APPENDIX F

### STUDY 5 MATERIALS

#### F.1 CONSENT FORM

##### STUDY OVERVIEW

In this study, you will be required to find an article related to your own current research topic, which you have never read before. Google is used as the search engine with a modified interface. You perhaps only occasionally use Google web search to find scholar articles. In this experiment we wish you to attempt this. Every search term you enter will be extended with “PDF” to favour articles as search returns. During the search task you will be provided with two lists of search results:

- *Keyword Search list*: the result list is generated from your search term.
- *Synonym Search list*: the first noun found in your search term is substituted by its synonym before searching. The result list is generated from this synonym search term.

In order to perform the task, you can search and browse web pages with no time limit. You will complete each session by submitting one selected article. You should submit only one article that interests you the most from any result list provided.

You will be required to rate each selected article and answer the questionnaire at the end of the experiment, every day. At the end of the final session you will be interviewed.

The experiment is divided into 5 sessions (one session per day). In each session you will be required to select one article, we expect you to spend up to roughly 30 minutes per session. For the participation you will be paid £20 cash at the end of the study.

### IMPORTANT INFORMATION

- Your recorded results will not be passed to any third party and are not being collected for commercial reasons.
- Participation in this study does not involve physical or mental risks outside of those encountered in everyday life.
- All procedures and information can be taken at face value and no deception is involved.
- You have the right to withdraw from the study at any time and to have any data about you destroyed. If you do decide to withdraw, please inform the experimenter.
- You will be reimbursed for your time upon successfully completing the study.

By signing this form you acknowledge that you have read the information given above and understand the terms and conditions of this study.

Age :	Gender :
Current Degree of study:	Department :
Level of Expertise in your research topic : Beginner / Intermediate / Advanced	
Signed :	Date :

Experimenter: Pawitra Chiravirakul, Dept. of Computer Science.PC345@BATH.AC.UK

## F.2 TASK MATERIALS

### BRIEFING SCRIPT

In this study, you will be required to find an article that relates to your research topic by using Google. During the search task you will be provided with two lists of search results; Keyword Search list and Synonym Search list.

This program will guide you through a training session before performing your search task.

In order to perform the task, you can search and browse web pages with no time limit. You will complete the task by submitting your selected article.

You will be required to rate your selected article and answer the questionnaire at the end.

---

### **TASK QUESTION**

- Based on your research topic, please select the article that interests you the most.

### **PRE-TASK QUESTIONS**

(Before starting session 2 - session 5)

- After the last session, have you done any literature search?
- Did you search on the same topic that you were working on in the last session?

Your last session topic: [participant's last session topic]

If Yes, please explain how you continued searching on that topic

- Did you find any useful literatures or ideas on the same topic after the last session? Please explain.